Combining cues for estimating slant in perception and action

van Mierlo, C.M.

2009

Cue combination for slant in perception and action

VRIJE UNIVERSITEIT

Combining cues for estimating slant in perception and action

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Bewegingswetenschappen
op woensdag 27 mei 2009 om 10.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Christa Mieke van Mierlo

geboren te Nijmegen

promotor: Prof.dr. J.B.J. Smeets
copromotor: dr. E.M. Brenner

Voor Rory en Findlay

Contents

Chapter 1: General Introduction

To pick up a cup of coffee and drink from it, we need to know the cup's shape, size, weight, roughness and flexibility of its material. We also need to know the distances our arm has to travel and the orientation our hand needs to be in for grabbing the cup, lifting it, and bringing it to our mouth. We estimate such attributes on the basis of information that our sensory organs receive. Even when restricting to a single sensory system, like the visual system, one can think of various instances in which the brain can use more than one source of information, from hereon referred to as *cue*, in order to estimate an attribute of an object. For example, we can estimate the distance of an object based on monocular cues such as size, height and texture density, but we can also exploit the difference in the object's projection in the two eyes, that is, the binocular disparity.

Optimal Cue Combination

Several studies have found that people combine such cues in a statistically optimal fashion (Ernst & Banks, 2002; Hillis *et al.*, 2004; Hogervorst & Brenner, 2004; Jacobs, 1999; Knill & Saunders, 2003; Landy *et al.*, 1995; Muller *et al.*, 2007; Oruc *et al.*, 2003; van Beers *et al.*, 1999). That is, subjects have been found to sum the estimates of different cues with weights that reflect their reliability in the current context as captured in the following equation:

$$P = \sum_i w_i c_i \qquad (1.1)$$

In equation 1, *P* represents our percept or representation of the attribute (which in the remainder of this thesis I will refer to as the combined estimate), w_i is the weight given to cue *i* and c_i is the estimate of cue *i*. For cues that are unbiased but contain independent Gaussian noise, a combination will be statically optimal if the weight (w_i) given to cue *i* depends on its precision to estimate the attribute (σ_i):

$$w_i = \frac{\sigma_i^{-2}}{\sum_i \sigma_i^{-2}} \qquad (1.2)$$

The lower the variability in a cue's estimate, the higher its weight in the combined estimate. This reliability of a cue, and hence its weight is thus thought to be primarily determined by its variability. This variability can either be the extent in which the cue's estimate varies over time, or the degree in which members of the same neuronal pool agree on the final estimate (the degree of noise in the pool, as suggested in Ma et al, 2006). Either way, the benefit of statistically optimal weighted averaging is that the combined estimate will be more precise than that of each of the individual cues.

Cue combination for slant

In my thesis, I studied different issues concerning how people combine cues for estimating slant. The slant of an object is its orientation in relation to the horizontal plane. Information about slant is available from monocular cues such as size, shape and texture density. That is, in the images projected onto the retina, the parts of the object that are further away will be smaller than the parts that are closer by. Their exact shape depends on the object's orientation, e.g., a circle's projection will be more elliptical if it is slanted towards or away from the viewer's body. The density of a texture increases gradually with distance. The depth information that is provided by binocular disparity will also give information about slant. Disparities will be bigger for parts that are closer by than for parts that are further away. By combining monocular and binocular disparities using statistically optimal linear weighted averaging (as described in Equation 1.1), one can obtain an estimate of slant that is more reliable than that of each individual cue. If only monocular cues and binocular disparities are used to estimate slant, we can express the weight given to the binocular disparities as 1 minus the weight given to the monocular cues. Equation 1.1 can then be rewritten as:

$$s_c = w_m s_m + (1 - w_m) s_b \tag{1.3}$$

In Equation 1.3, S_c represents the combined estimate of slant, s_m represents the slant estimate provided by the monocular cues and s_b represents the slant estimate provided by the binocular disparities. Several studies have indeed found evidence that slant estimates from monocular and binocular cues are optimally combined under constant conditions (Hillis et al., 2004; Knill & Saunders, 2003; Muller et al., 2007).

Scope of this thesis

I investigated two issues in optimal cue combination. In Chapters 2 and 3, I explored how the timing of different cues affects their combination into one optimal estimate, in perception (Chapter 2) and perception & action (Chapter 3). In Chapter 4, I studied whether the weight of a cue is indeed solely determined by its variability, even if subjects perceive its estimate to be incorrect. In Chapter 5, I studied how stereoblind people can perceive motion in depth from a difference in motion between the eyes.

Chapters 2 and 3: Timing

FMRI Studies have found that shape from monocular cues such as contour, shading and perspective activate the lateral and ventral occipital cortex (LOC) (Kourtzi & Kanwisher, 2000). Topographical representations of shape from binocular disparity have been found in MT/V5, LO-1 and LO-2 within area KO, also the ventral (LOC) visual areas, and areas V3 and hV4 (Bridge & Parker, 2007; Chandrasekaran *et al.*, 2007). Though shape from monocular cues and disparity seem to share at least one area in their processing (LOC), their independent processing stages might result in differently timed estimates (as shown for other independently processed cues that are related to the same stimulus by Schmolesky et al., 1998). Such timing differences might not be important when we only use one of the cues, but when we combine them in one estimate, single-cue estimates that are associated with different times might erroneously be integrated. This might be especially harmful when we have to respond fast and cannot wait for new information to become available, as when adapting our ongoing motion to new information about the planned end position of our movement.

In Chapters 2 and 3, I investigated how latency differences between cues affect their combination in one estimate, in both perception and action. I introduced artificial timing differences between the moments monocular and binocular cues indicated a changed slant and studied how this affected subjects' detection and discrimination of slant changes (Chapter 2) and fast manual responses (Chapter 3). Are timing differences between slant cues simply ignored when we combine them in one estimate or do we somehow compensate for asynchrony?

Chapter 4: Precision and bias

The perceptual information that we use to plan and control action not only needs to be precise, but also needs to be correct. In the real world some cues might not give a correct estimate for the attribute in question, as for example we are looking at a distorted reflection of an object, instead of the object itself. Combining such cues using Equations 1.1 and 1.2 will result in combined estimates that will be precise, but not correct. Does the brain, when it knows that some cues are biased, really only take the precision of these cues' estimates into account?

Several studies have found that subjects increase the weight of a cue that is consistent with (haptic or auditory) feedback and decrease the weight of a cue that is not (Atkins *et al.*, 2001; Ernst *et al.*, 2000; Jacobs & Fine, 1999). In all of these studies, the physical precision of the cues was not manipulated; only the cues' perceived bias varied from session to session. This suggests the perceived correctness of a cue might also influence its weight in the combination rule.

However, the physical precision of a cue might not be the same as its perceived precision, as the perceived precision is not only influenced by the incoming sensory information but also by the person's expectations about the world. For example, in a study by Knill (2007), subjects changed the weights that they gave to foreshortening and binocular disparity when large conflicts between the cues indicated that the as circles perceived stimuli might need to be reinterpreted as being ellipses. Knill argued that a broadening of shape expectations led to a decrease in judged reliability and thus the weight that was given to shape related cues. In Knill's study, the change in weights took place slowly as subjects got more experience with the conflict. In the feedback studies (Atkins *et al.*, 2001; Ernst *et al.*, 2000; Jacobs & Fine, 1999) the change in weights was also slow. Might the change in weights found in these studies not be related to a change in the judged correctness of the cues, but to a change in their judged precision? All these studies used a cue conflict paradigm to measure the weights that were given to the different cues. When generating such conflicts, the shape of the stimulus on the screen gets slightly distorted when compared to its shape on non-conflict trials. A decrease in the judged precision of shape related cues might therefore not be unthinkable, although this does not explain the decrease in the weight given binocular disparity when this cue is indicated to be incorrect (since binocular disparity can be independently determined from monocular shape). In Chapter 4, I therefore investigated whether the change in weights that is seen in response to feedback about the correctness of the cues, results from a change in their perceived precision or from a change in their perceived correctness. In this last case, not only the variability of a cue influences its weight, but also its bias. This cannot be accommodated by current optimal cue combination theory.

Chapter 5: Motion in depth

Chapter 5 stands rather apart from the rest of the thesis. The topic of this study arose by accidental discovery when I was performing control experiments for Chapter 3. JS is stereoblind as a consequence of the eye patch treatment that he received to treat amblyopia in his childhood. He participated in many of my experiments to check for monocular artifacts that could be used to respond instead of my manipulations of binocular disparity. Surprisingly, in the experiment described in Chapter 3, he could respond to the changes in binocular disparity. When he performed the experiment monocularly, his responses were eliminated. This indicates that JS can use some sort of difference in motion between the two eyes to determine the range of slants in which the surface is likely to be in at the end of its movement.

Several studies have found that subjects with normal binocular vision, as well as subjects with strabismus (who have low stereo-acuity), can use interocular velocity differences (IOVDs) to judge motion in depth (Brooks & Stone, 2006; Fernandez & Farell, 2006; Kitaoji & Toyama, 1987; Maeda *et al.*, 1999; Rokers *et al.*, 2008). The fact that people with low stereo-acuity can use this cue suggests that the pairing of velocities between the eyes for calculation of this cue might not be based on position. Therefore, JS might use IOVDs to determine the direction in which the surface was moving in Chapter 3. In Chapter 5, I explored whether three stereoblind subjects can use IOVDs to perceive motion in depth, and if so, on the basis of what information they correspond the motion between the two eyes for calculation of this cue.

Chapter 2:
Temporal Aspects of Cue Combination

Abstract
The human brain processes different kinds of information (or cues) independently with different neural latencies. How does the brain deal with these differences in neural latency when it combines cues into one estimate? To find out, we introduced artificial asynchronies between the moments that monocular and binocular cues indicated that the slant of a surface had suddenly changed. Subjects had to detect changes in slant, or to indicate their direction. We found that the cues were combined to improve performance even when the artificial asynchrony between them was about 100 ms. We conclude that neural latency differences of tens of milliseconds between cues are irrelevant because of the low temporal resolution of neural processing.

Introduction
People use various kinds of information to make sense of the visual input from the external world. For example, they estimate the slant or orientation of a surface from texture gradients, motion parallax, retinal shape, binocular disparity, and so on. The brain is believed to process various kinds of information (cues) in different visual pathways in the brain, with neural latencies that can differ by tens of milliseconds (Schmolesky et al., 1998). After such independent processing, the brain combines different cues for the same property into a single estimate that is more reliable than any of the estimates based on the individual cues (weighted averaging; Ernst & Banks, 2002; Hillis et al., 2004; Jacobs, 1999; Knill & Saunders, 2003; Landy et al., 1995; van Beers et al., 1999). The contribution of each cue to this estimate is thought to primarily be determined by its reliability, but it could also be influenced by other factors such as the likelihood of the value indicated by the estimate occurring, the consistency between different cues, or the correlation between the errors of the two cues (Hogervorst & Brenner, 2004; Knill & Saunders, 2003; Landy et al., 1995; Oruc et al., 2003).

Differences in neural latency between cues about unrelated attributes (such as colour and motion) may be responsible for the large systematic errors that subjects make when trying to synchronise changes within such cues (Arnold & Clifford, 2002; Aymoz & Viviani, 2004; Moutoussis & Zeki, 1997a, 1997b; Nishida & Johnston, 2002; Viviani & Aymoz, 2001; Wu et al., 2004; Zeki & Moutoussis, 1997). But differences in neural latency need not only occur for cues that provide information about unrelated attributes. Cues that describe the same property or attribute of a stimulus are also likely to have different neural latencies (Greenwald et al., 2005). Do these timing differences lead to systematic errors when such cues are combined, or are there special mechanisms in the brain for preventing this?

Aymoz and Viviani (2004) found that people tend to make smaller systematic errors in synchronizing changes in color and movement when the changes were the consequence of another person's actions. They speculated that the action of the other person activated a specialized system that re-establishes synchrony within the brain by compensating for the neural delay between the cues. Bartels and Zeki (2006) showed that people were better at synchronizing cues that describe the same attribute than at synchronizing cues that describe different attributes. So there might be special mechanisms for dealing with timing differences between cues that are normally combined, such as cues for the same property.

Several cross-modal studies suggest that precise synchrony of different cues might be irrelevant due to the relatively low temporal resolution of neural processing. Munhall et al. (1996) showed that the McGurk-effect is robust for lags of up to 180 ms. Shams et al. (2002) found that subjects perceive a single visual flash as two flashes when it is accompanied by two auditory beeps. Similarly, auditory sequences of beeps have been found to modulate the tactile perception of sequences of taps (Bresciani et al., 2005). These effects persisted even when the flashes, beeps and taps were separated by more than 100 ms. So delays of up to 100 ms seem to be tolerated when integrating cues across modalities. Is this also the case for cues for the same attribute within a single modality? In particular, can the benefits of combining cues (e.g. through weighted averaging) be obtained without precise temporal synchrony?

Experiment 1

We conducted an experiment in which we explored the sensitivity of cue combination for asynchronies between binocular and monocular slant cues. Subjects had to detect changes in the slant of a plane. The orientation of the plane was evident from binocular disparity and from monocular information. The binocular cue could vary independently of the monocular cue, so that either one cue or both cues could indicate a change in slant. When both cues indicated a change, the timing of the change could differ for the two cues, thus creating different artificial cue asynchronies. We studied how subjects' detection of the changes in slant varied as a consequence of the asynchrony between the cues. To determine whether subjects were really combining the cues for detecting the change, we compared their performance when both cues changed, with the performance predicted by probability summation on the basis of their performance when only the binocular or only the monocular cue changed (Hillis et al., 2002; Poom, 2002; Wuerger et al., 2003).

Methods

Subjects

Six subjects participated in the experiment, three female and three male. One subject was an author; the other five were volunteers who were naïve about the purpose of the experiment. All subjects had normal binocular vision; their stereo acuity was better than 60 arc seconds (tested with RandotTM plates).

Apparatus and stimuli

A Silicon Graphics Onyx Reality Engine was used to present the stimuli on a CRT monitor (120 Hz; horizontal size: 39.2 cm, 815 pixels; vertical size: 29.3 cm, 611 pixels; spatial resolution refined with anti-aliasing techniques). The subject sat 40 cm from the monitor, resting his/her head on a chin rest. The subject was wearing liquid crystal shutter spectacles that successively blocked each eye in synchrony with the refresh rate of the monitor (120 Hz), so that different images were shown to the left and right eye in rapid alternation. A new image was presented to each eye every 16.7 ms (60 Hz). The individual's inter-ocular distance was taken into account when creating the image presented to each eye. As a result, both the subject's ocular convergence and the retinal images were appropriate for the stimulus at the simulated distance.

Stimuli

The stimuli were designed so that monocular and binocular information about changes in slant could be manipulated independently. The stimulus was a simulated red ring within which 10 dots were randomly distributed. The ring had an outer radius of 70 mm, and a width of 10.5 mm. The dots had a diameter of 5 pixels. The dots were added to increase the strength of the binocular disparity cue. There were so few dots within the ring that the contribution of their distribution to the monocular estimate of slant was probably negligible. Every 16.7 ms the ring changed its position to a new random position within 20 mm of the centre of the simulated surface (which coincided with the centre of the screen). At the same time, new random positions were chosen for the dots. This prevented subjects from detecting slant changes on the basis of any motion in the image. The ring and dots were presented to the left eye first and then to the right eye, before being replaced by a similar ring and dots at a slightly different position. Subjects perceived this stimulus as several rings that jittered on a plane. The slant of this plane was defined by the binocular disparities, the shape of the ring, and the distribution of the dots.

Since the ring's shape and dots' distribution always indicated the same slant, we will refer to them together as a monocular cue. In order to change the slant indicated by the monocular and binocular cues independently, we determined how a surface with a slant defined by the monocular cue would look for a single (cyclopean) eye, and then rendered images for the two eyes that on average provide this retinal image, while having the binocular slant that we wanted (Knill, 1998; Landy et al., 1995).

Most of the time, the binocular and monocular cues both suggested that the slant of the plane in which the rings seemed to jitter was 10° (base slant), with a positive angle meaning that the top was further away than the bottom. The slant could increase abruptly by 5°, 10°, 15°, 20° or 25°, meaning that the top always tilted further away. These slant changes could occur in the binocular cue alone, in the monocular cue alone, or in both cues. When both cues changed their slant, they could do this simultaneously, or asynchronously with one of 8 different timings: the change in the binocular disparity cue could occur 400, 200, 100 or 50 ms before or after the change in the monocular cue. The next slant change occurred between 4 and 6 s after the slant had returned to its baseline value for both cues. The plane regained its base slant gradually within 400 ms; it returned slowly so that the subjects would not perceive this as a second change (see Figure 1.1).

Procedure

Subjects saw a set of rings jittering on a plane that occasionally changed its slant. They had to respond to any change in slant by pressing the right mouse button. No feedback was given. In total there were 55 conditions. For each of the 5 amplitudes of slant change there were 9 two-cue conditions with various asynchronies (including the 0 ms asynchrony) and two single-cue conditions. Subjects performed 20 trials per condition, so 1100 trials in total, distributed over several sessions. The slant change for each trial was randomly selected from these 55 conditions.

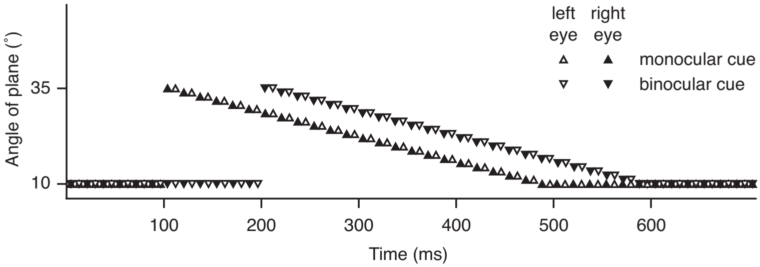

Figure 1.1. Schematic representation of one of the conditions of experiment 1 in which both cues change by 25°, with a 100 ms delay between the changes. Each frame on the screen is represented by two symbols: one indicating the value of the monocular cue (the upward pointing triangles) and another indicating the value of the binocular cue (the downward pointing triangles).

Data Analysis

We considered subjects to have detected the change if they responded between 150 ms after the first cue changed and 1 s after the last cue changed (response interval). We determined the fraction of slant changes detected for each cue asynchrony. If the binocular and monocular cues are processed completely independently of one another, and independently give rise to responses within the allocated time, the asynchrony between them should be irrelevant, and the probability of detecting a slant change when both cues change (P_{both}) is the chance of not missing the slant change in both cues, which can be calculated on the basis of the subject's performance for the single-cue conditions ($P_{binocular}$ and $P_{monocular}$):

$$P_{both} = 1 - (1 - P_{binocular})(1 - P_{monocular}) \qquad (2.1)$$

If we find that performance for a particular two-cue condition is better than predicted by probability summation (P_{both} in Equation 2.1), we can conclude that subjects detected the corresponding slant changes better than was to be expected on the basis of simply having two chances to react. Such better performance would imply that the cues must have been combined in a clever manner that makes identifying changes in slant more reliable. If such better performance is found, then a comparison of the different cue asynchronies will reveal the temporal sensitivity of the underlying cue combination.

For each cue asynchrony and amplitude, we used a paired t-test to examine whether performance was better than predicted by probability summation (Equation 2.1). We also tested with a paired t-test per cue asynchrony whether performance was poorer for the asynchronous than for the synchronous two-cue condition. Using t-tests in this manner is a conservative way of determining whether performance differs between the conditions, because the benefit that we can expect from cue combination depends on the relative resolution of the two cues, which is likely to differ between subjects. If we find that performance on the two-cue conditions is systematically better than predicted by probability summation and depends on the timing difference between the cues, we will have a strong indication that subjects combined the two cues into one estimate of the change in slant. Differences between the asynchronies will then reveal the temporal resolution of combining the cues.

We estimated uncertainty bounds (standard error of the mean of the binomial distribution) for each condition and subject from the observed fraction of slant changes detected (P) and the number of samples in each condition (N = 20) :

$$\mathrm{SEM} = \frac{\sqrt{\mathrm{P}(1 - \mathrm{P})}}{\sqrt{\mathrm{N}}} \qquad (2.2)$$

These values were averaged across subjects and conditions to give a single estimate of the within-subjects standard error for each amplitude of change in angle.

Results

Our subjects' average performance is displayed in Figure 2.2. On average, subjects detected changes in the binocular single cue condition better than changes in the monocular single cue condition (compare the upward and downward triangles in each panel). Subjects responded to between 39 and 56% of the changes in the binocular cue, and to between 8 and 55% of the changes in the monocular cue. For slant changes in both cues, only detection of 20º slant changes with an asynchrony of +50 ms between the cues was significantly better than predicted by probability summation ($t(5) = 3.08$, $p < .05$). Even if there is no benefit from cue combination, one would expect more than one out of 45 comparisons to appear to be significant at the 5% level.

For the small changes in slant (amplitudes of 5° and 10°) performance seemed to be systematically worse than predicted by probability summation. For the larger changes in slant (15°, 20° and 25°) a broad performance peak around the smaller cue asynchronies was apparent. The 20° slant changes with an asynchrony of -400 ms between the cues and the 25° slant changes with either a +100 or a +400 ms asynchrony between the cues were significantly less likely to be detected than synchronous slant changes of the same amplitude (t (5) > -2.02, $p < 0.05$). Three significant comparisons out of forty is only one more than one would expect by chance alone.

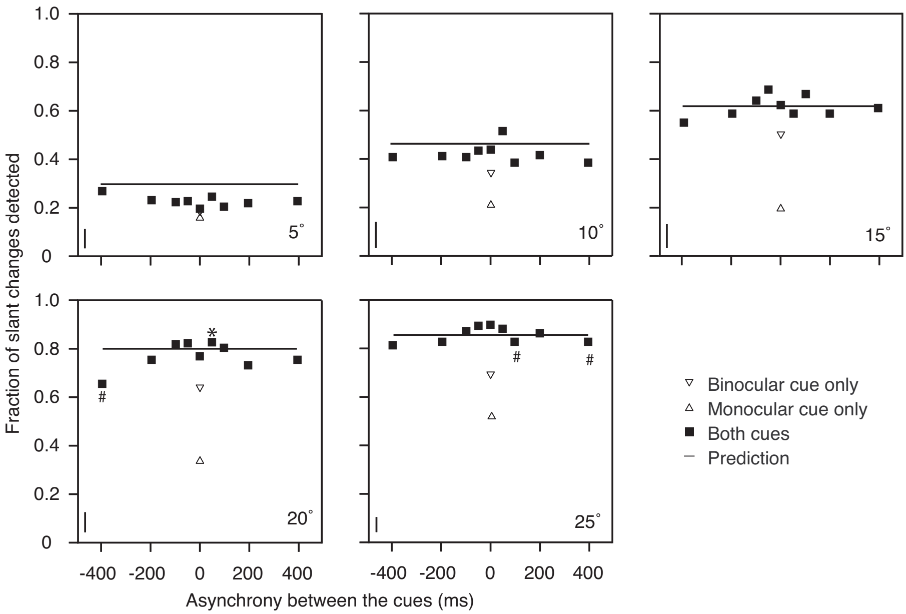

*Figure 2. Average performance in Experiment 1. Each panel gives the results for one of the five different amplitudes of change. Positive values of the asynchrony indicate that the monocular cue changed after the binocular cue. The data for changes in a single cue are plotted at an asynchrony of zero. The error bar at the bottom left of each graph is an estimate of the within subjects standard error for the two-cue performance (averaged across asynchronies). A * indicates significantly better performance than predicted by probability summation. A # indicates significantly worse performance than for the 0 ms asynchrony.*

Discussion

Performance was significantly better than probability summation for only one amplitude-asynchrony condition. Performance often even appeared to be worse than predicted by Equation 2.1. A possible reason for this might be that Equation 2.1 does not consider false positives: correct responses that are independent of the actual change. In our analysis we assumed that all responses that people made resulted from them really detecting the change. However, subjects sometimes seemed to misinterpret the jitter in the ring as a change in slant. We know that such false positive responses occurred because we regularly observed responses long (>2 s) after the change had occurred. Presumably, such responses also occur when a change has taken place but was not detected. Since people are likely to make as many false positive responses in the two-cue conditions as in each single-cue condition, Equation 2.1 will overestimate the predicted performance for the two-cue conditions because it incorporates the false positive responses twice: once in $P_{binocular}$ and once in $P_{monocular}$. The conditions with larger asynchronies may contain slightly more false positives than the synchronous conditions because of their longer response intervals. Moreover, people are more likely to respond when they do not detect the target, because once they have detected it there will temporarily be no need to respond, so the number of false positives is likely to depend on the subject's performance. We therefore propose that the predictions in Figure 2.2 lie higher than they should, as a result of ignoring false positive responses in Equation 2.1. This would explain why performance was no better than predicted by probability summation despite the apparent peak at small cue asynchronies for the larger amplitudes of slant change.

Experiment 2

If our impression that there is a peak in performance for small cue asynchronies is correct, the peak's width suggests that an asynchrony between the cues of up to 100 ms hardly influences the benefit that is obtained from combining the cues. However, this proposal rests on the assumption that we overestimated two-cue performance in experiment 1 as a consequence of not accounting for false positives. In our second experiment we therefore asked subjects to perform a task that allowed us to determine the number of false positives that they make. They now had to indicate the direction of any slant change that they saw. Since subjects could respond both incorrectly and correctly when not responding to an actual change in slant, the term "false positives" no longer adequately describes such responses. We will refer to these responses as "guesses" from now on. If subjects guess, about half of their guesses will be correct and half of them will be incorrect. The number of guesses is therefore twice the number of errors in indicating the direction of the change. By removing subjects' guesses from their responses before applying Equation 2.1, we can calculate predictions for probability summation in which guesses are considered.

Methods

Subjects

The same six subjects participated in the second experiment.

Apparatus, Stimuli and Procedure

We used the same set-up as in the previous experiment, but made a few changes to the stimuli and procedure. In the previous experiment, our subjects detected changes in binocular disparity more easily than changes in the monocular cue. We therefore used a larger base slant (25º relative to frontal) to increase the reliability of the monocular cue (Knill, 1998). The change in slant always had an amplitude of 20º, but it could be in either direction. The combination of a 25º base slant and a ± 20º change ensured that the surface never crossed the fronto-parallel plane, which is important because doing so could make the changes in the monocular cue ambiguous, or at least less clear. As in the previous experiment, slant changes could occur in binocular disparity, in the monocular cue, or in both, with time intervals ranging up to 400 ms. Subjects were instructed to indicate the direction of any slant changes that they detected. They pressed the left mouse button for slant changes "backwards" and the right mouse button for slant changes "forwards" (with the direction referring to the movement of the top of the surface).

Data Analysis

Since choice reaction times are known to be longer than simple reaction times, we gave subjects slightly more time to respond. We determined the fraction of detected slant changes (both incorrect and correct responses) within an interval starting from 150 ms after the change in the first cue up to 1.2 s after the change in the last cue (when only one cue changed it was both the first and the last). We assume that the responses that the subjects make consist of a number of real detections and a number of guesses (including responses to changes other than the simulated slant changes). Since these guesses are as likely to be correct as incorrect, we assume that the number of guesses is twice the number of incorrect responses. Equation 2.1 only applies to the number of slant changes that subjects detected, not to their guesses. So before we use this equation to predict two-cue performance, the fraction of guessed responses has to be removed from all the P values. The fraction of changes that were detected (P_d) can be estimated from the total fraction of trials in which the subject responded (P_r) and the fraction in which the subject responded incorrectly (P_e):

$$P_d = P_r - 2P_e \qquad (2.3)$$

Equation 2.3 holds independently for each condition, so substituting P_d for the P's in Equation 2.1 gives:

$$P_{r_{both}} - 2P_{e_{both}} = 1 - (1 - (P_{r_{monocular}} - 2P_{e_{monocular}}))(1 - (P_{r_{binocular}} - 2P_{e_{binocular}})) \quad (2.4)$$

Equation 2.4 can be used to take guesses into account when predicting the fraction of responses for presentations with two cues ($P_{r_{both}}$) on the basis of single-cue performance ($P_{r_{monocular}}$ and $P_{r_{binocular}}$).

For each of the 9 two-cue conditions we used paired t-tests to examine whether the observed two-cue performance was significantly better than the value predicted on the basis of Equation 2.4. We also used 8 paired t-tests to examine whether performance for each asynchrony was poorer than that for the synchronous slant changes.

Results

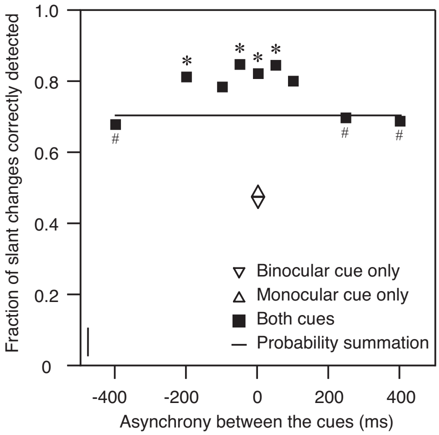

*Figure 3. Average performance in Experiment 2. A * indicates significantly better performance than predicted by probability summation. A # indicates significantly worse performance than for the 0 ms asynchrony. Other details as in Figure 2.*

We first determined the fraction of incorrect responses (P_e) for each subject and each condition. An analysis of variance on P_e with Condition and Subject as factors revealed that P_e did not differ significantly between the different conditions (F (10) = .784, p = .644). Since it is important to get a reliable estimate of P_e, and the number of guesses is quite modest, we determined a single value for each subject and used this value for $P_{e_{both}}$, $P_{e_{monocular}}$, and $P_{e_{binocular}}$ in Equation 2.3.

Figure 3 shows the average fraction of slant changes that was detected. Subjects responded to 65.6% of the forward slant changes and 78.7% of the backward slant changes. For four of the nine asynchronies (-200, -50, 0, 50 ms), the paired t-tests revealed that performance was better than predicted by equation 3 ($t(5) > 2.038$, $p < .05$). Performance for the –400, +200 and +400 ms cue asynchronies was significantly poorer than for the synchronous condition ($t(5) > 2.276$, $p < .05$).

Discussion

For asynchronies up to about 100 ms, performance with both cues was clearly better than predicted on the basis of probability summation. Probability summation did reliably predict performance for the largest cue asynchronies (±400 ms), confirming that our analysis now addresses all major issues, because as the asynchrony is increased the cues must at some moment become independent. So the findings of experiment 2 support the suggestion from experiment 1 that subjects combine the two cues even when there are small timing differences between them. Additionally, there appears to be a shift in the optimal delay toward negative asynchronies, which is consistent with binocular slant cues being processed faster than monocular slant cues (Greenwald et al., 2005). As already mentioned, cross-modal cue combination is known to persist with asynchronies of slightly more than 100 ms (Munhall et al., 1996; Shams et al., 2002). The similar tolerance of asynchronies in the present study shows that in this respect cues for the same property (slant) within a single modality (vision) are not treated in a special manner.

Experiment 3

In the third experiment we examined the validity of two assumptions that we made when interpreting the data of experiments 1 and 2. The first is that the lack of change in one cue does not influence the detection of a slant change in the other cue. The second is that binocular disparities and retinal shape do not interact before providing estimates of slant. Further assumptions are discussed in the general discussion.

Cue conflict in single-cue conditions

In the single cue conditions, one cue remained in base slant while the other changed its slant, thus creating a cue conflict during the slant change. Our analysis was based on the assumption that for these conditions, the detection of a slant change in one cue was not affected by the unchanging slant of the other cue. Based on this assumption we concluded from the findings of experiment 2 that detection of slant changes is better than predicted by probability summation when the two cues change in close temporal proximity. In the third experiment we compared performance in a replication of the single-cue conditions of experiment 2 with performance in two new single-cue conditions in which the slant conflict during the slant change was reduced. If performance in the new conditions turns out to be better than in the original conditions, we would have to consider the possibility that performance in the previous experiments was not better when changes in two cues were combined, but worse when one cue indicated that there was no change.

Independency of processing

Up till now, we have assumed that the two cues are processed independently, before the brain combines them into one estimate of slant. Tittle and Braunstein (1993) suggested that this assumption might not hold for all cues within the visual system. For shape judgments from binocular disparity and motion parallax, they found that the presence of motion in a stereo display helped solve the binocular-correspondence problem, so that motion helped establish the binocular estimate of shape as well as providing an independent estimate of shape. Adams & Mamassian (2004) showed that texture information can also modulate shape from disparity in a way that is inconsistent with simple linear cue combination.

We investigated whether the cues in our experiment interacted before they each supplied an independent estimate of slant with the help of two new two-cue conditions. In the first, the two cues signaled a slant change on alternate pairs of frames in rapid sequence (a pair of frames means one frame per eye). In the second, the two cues signaled the change simultaneously once every two pairs of frames (see Figure 2.4). In both cases each cue alternates rapidly between the new slant and the base slant, but in the asynchronous condition the two cues are always in conflict, whereas in the synchronous condition the cues always agree. If the cues do not interact before providing estimates of slant (and the cue combination process is not very sensitive to the precise timing of the estimates, as we have already seen) then performance in the two conditions should be the same. If we find better performance when the two cues change simultaneously we would have evidence that the cues interact before they each generate an estimate of slant.

Methods

Subjects

Nine subjects participated in this experiment: five female and four male. Six of the subjects had also participated in the former two experiments. All subjects had normal binocular vision; their stereo acuity was better than 60 arc seconds (tested with Randot™ plates).

Apparatus, Stimuli and Procedure

The same set-up was used as in the former two experiments. We repeated the single-cue conditions from the second experiment for our modified procedure (see below), and added four new conditions. In a new monocular single-cue condition, the change in the monocular cue was only presented to one of the eyes (by simply not drawing the images for the other eye) so that there was no conflicting binocular disparity cue. In a new binocular single-cue condition, only the dots that were previously used to fill the ring were visible. Omitting the ring practically eliminated the monocular cue, so that the conflict was very much reduced, while leaving the binocular cue largely intact.

We introduced two additional conditions: an asynchronous and a synchronous two-cue condition. In the asynchronous condition, the slant changes were specified by both cues in rapid sequence. That is, in one pair of frames the monocular cue specifies base slant while the binocular cue indicates a changed slant, while in the next pair of frames the monocular cue indicates a changed slant and the binocular cue specifies base slant. In the synchronous two-cue condition, both cues specify a changed slant simultaneously once every two pairs of frames (see Figure 2.4), both specifying the base slant during the other pair of frames.

As in experiment 2, the base slant of the surface was 25°, and the ring could change its slant by ±20°. To simplify the analysis, we changed our paradigm to a Forced Choice procedure. Our subjects had to indicate the direction of the slant change after an auditory signal indicated that a slant change had occurred.

Data Analysis

Due to the simplified procedure we could just compare the proportion of correct responses between the different conditions. We used one-sided Chi-square tests to examine whether there were more correct responses in the new binocular single-cue condition than in the original binocular single-cue condition, more in the new monocular single-cue condition than in the original monocular single-cue condition, and more in the synchronous than in the asynchronous two-cue condition.

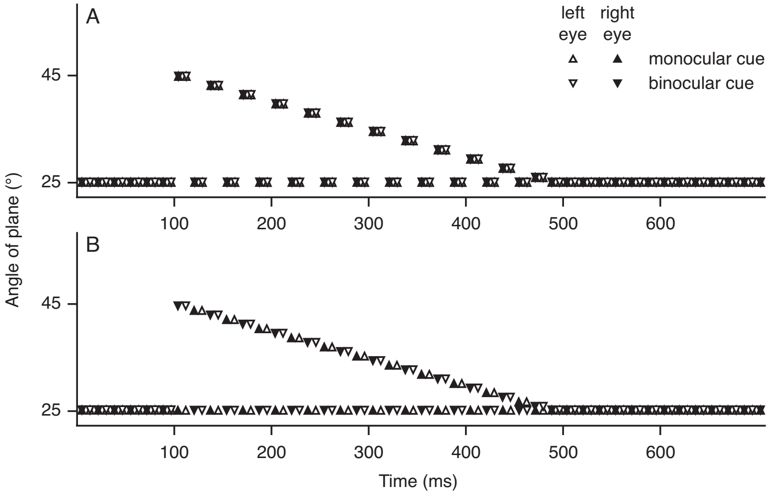

Figure 4. Schematic representation of the synchronous (A) and the asynchronous (B) two-cue conditions of Experiment 3. Both panels show a 20° slant change from a 25° base slant. In the synchronous condition the cues are never in conflict, whereas in the asynchronous condition they are in conflict whenever the surface is not at the base slant.

Results

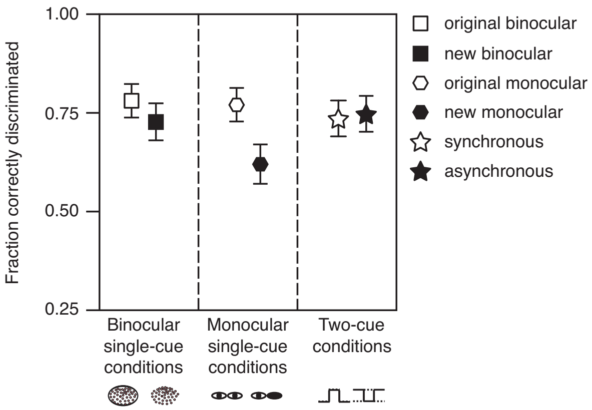

Figure 5. Average performance in Experiment 3. Error bars show 95% confidence intervals across subjects.

Our subjects' average performance is displayed in Figure 2.5. None of the 3 Chi-Square tests showed significant improvements. The average performance in the new single-cue conditions even seems to be worse than in the corresponding original single-cue conditions.

Discussion

Reducing the cue conflict in the single-cue conditions did not improve performance. It even decreased the number of correct responses that subjects made, especially in the monocular single-cue condition. This is probably because presenting the slant change to only one of the eyes doubled the interval between the frames in which it was present. Similarly, removing the ring in the new binocular single-cue condition reduced the amount of binocular information, as well as practically eliminating the monocular information. Thus subjects had slightly less of the relevant information available in the new single-cue conditions than in the original single-cue (and two-cue) conditions of experiment 1 and 2. This questions the validity of our control experiment to some extent, but it is clear that the benefits of not having a cue that does not change (while the other does) do not outweigh the costs of reducing the amount of information in the new single cue conditions. Thus the lack of slant change in one of the cues cannot have much effect on how readily a change in the other cue is detected. Subjects' performance in the new asynchronous condition was no worse than in the new synchronous condition. This finding indicates that there was no interaction between the cues before each provided an estimate of the slant change. From experiment 3 we can conclude that there is no need to reconsider our interpretation of experiment 2.

General Discussion

In the present study we examined how the detection of changes in surface slant was affected by artificial delays between binocular and monocular cues. We found a benefit for detecting two-cue slant changes beyond that predicted by probability summation, even when the two cues changed at moments that differed by tens of milliseconds. This implies that neural latency differences between visual cues will seldom be an issue for the brain when it combines cues into one estimate. Apparently, either the processing of the cues themselves or their combination into one estimate has quite a poor temporal resolution. However, this conclusion rests on one final assumption that needs to be discussed.

We assumed throughout the study that the cues are used independently to detect changes in slant, and that if there is evidence from more than one cue that the slant of the surface in which the ring jittered changed, this evidence is combined to obtain a more reliable estimate of the change. Only considering information about changes in slant seems reasonable to us because the visual system is generally most sensitive to transients. However, there are alternative ways to combine the information provided by the two cues. Evidence for a change in one cue might be combined with evidence for no change in the other cue when the cues do not change simultaneously (for evidence against this option see experiment 3). It is also possible that the cues are continuously combined to give a single estimate of slant, and subjects detect changes in this combined estimate. Finally, subjects might only notice a difference between the changed combined estimate of slant and the baseline value. Would any of these three alternatives influence our conclusion that the temporal resolution of processing and combining changes in visual slant cues is poor?

If evidence for a change in one cue is combined with evidence for a lack of change from the other cue, the first of these two changes would be equivalent to a change in a single cue. When the second cue changes, the value of the first cue is different than in the single-cue conditions, but the change of the cue in question is identical to the change of that cue alone. The fact that the first cue is slowly changing back to its original baseline value might even slightly decrease the probability of detecting the change, because it is in the opposite direction than the change that is to be detected. So according to the first alternative there is no reason to expect performance to be any better than predicted by probability summation. Our finding of an increased probability of detecting the slant changes when the two cues changed within tens of milliseconds of each other can only be consistent with this alternative if the changes are considered to overlap in time. experiment 3 provides additional evidence against this alternative.

The second alternative is that subjects detect changes in a combined estimate of slant. If so, asynchronous slant changes would be equivalent to two smaller slant changes. The first change is identical to a change in a single cue. However, the second change is not because it starts from a different perceived slant. It is not evident that this should make the changes easier to detect, because the change within the single relevant cue is exactly equivalent, even if the perceived initial and final slants of the second change are different. Moreover, if the perceived slant at the time of the second change influences performance, why should detection improve in a similar manner both when the perceived slant is higher than the baseline value of 10º in experiment 1, and when it is sometimes higher and sometimes lower than 25º in experiment 2? Again, the simplest explanation would be that the large range of asynchronies for which performance is better than probability summation arises because the temporal resolution of the processing underlying the judgment of slant is so low that the two changes merge into one larger slant change.

The third alternative is that subjects do not respond to transients (changes in the perceived slant) at all, but sometimes notice that the slant is no longer at the baseline value. Since each cue's slant changed back gradually, the combined value would also change gradually, which could explain why subjects' performance was quite insensitive to timing differences between changes in the two cues without having to rely on a poor temporal resolution of visual processing or cue combination. We do not find this very likely because the visual system is generally most sensitive to transients, the task was to detect slant changes, and none of the subjects ever reported seeing two changes in rapid sequence. However, our findings do not rule out this possibility. In particular, performance for approximately synchronous changes in both cues is not evidently or systematically different from performance for twice the amplitude of the change in a single cue (see Figure 2.2). Thus if subjects detect the change by noticing that the slant is no longer at baseline, rather than noticing the change itself, we cannot be certain from our study that the processes involved have a low temporal resolution. However, such a mechanism would make timing differences between cues much less relevant an issue, because a temporal error of tens of milliseconds only introduces large differences between cues at moments at which the image suddenly changes, such as when one makes saccades or when there are quickly moving objects. Thus asynchronies might be tolerated because they are only present for short periods of time.

We conclude that timing differences between cues are unlikely to be an important issue in human slant perception. Probably the temporal resolution of the processing of individual cues is so low that differences in timing can be ignored when combining cues. In daily life, external events will always cause cues to change in synchrony, so differences in timing between visual cues within the brain only arise from differences in neural processing time. Since the reported differences in visual processing time (Schmolesky et al., 1998) are modest when compared with the low temporal resolution that we find for cue combination, it is unlikely that these differences in processing time have much influence on our perceptual judgments or need to be compensated for.

Chapter 3:
Slant cues are processed with different latencies for the online control of movement

Abstract

For the online control of movement it is important to respond fast. The extent to which cues are effective in guiding our actions might therefore depend on how quickly they provide new information. We compared the latency to alter a movement when monocular and binocular cues indicated that the surface slant had changed. We found that subjects adjusted their movement in response to three types of information: information about the new slant from the monocular image, information about the new slant from binocular disparity, and information about the change in slant from the change in the monocular image. Responses to changes in the monocular image were approximately 40 ms faster than responses to a new slant estimate from binocular disparity, and about 90 ms faster than responses to a new slant estimate from the monocular image. Considering these delays, adjustments of ongoing movements to changes in slant will usually be initiated by changes in the monocular image. The response will later be refined on the basis of combined binocular and monocular estimates of slant.

Introduction

When we want to place an object on a surface, we need to estimate the surface's slant to make sure that the object has about the same orientation as the surface before making contact. Information about this orientation is available from binocular disparity and from the monocular images. The information in the monocular images includes cues such as the shape of the surface's projection on the retina, changes in texture density across the retina, and motion parallax. Different slant cues are likely to be processed at different rates, and so may provide information about changes at different latencies. Previous research suggested that differences in latency are ignored, so that cues with shorter latencies influence the combined estimate earlier (Van Mierlo et al., 2007).

One way to examine how people use visual information to guide their action is by examining how they respond to perturbations of such information during their movement (Brenner & Smeets, 1997; Goodale et al., 1986; Saunders & Knill, 2003; Veerman et al., 2008). Different studies have reported different latency differences between monocular and binocular cues. In a study in which subjects had to respond to perturbations in surface slant (Greenwald et al., 2005), slant estimates based on binocular disparity appeared to influence corrections earlier than slant estimates based on monocular cues, so the authors concluded that binocular disparity was processed more quickly. This finding is surprising, because Allison & Howard (2000b) found that perceived slant shifted from being dominated by perspective to being dominated by disparity as exposure time to a test stimulus increased. Moreover, Brenner and Smeets (2006) found that subjects corrected movements faster in response to a jump in target depth when the jump was visible as a change in the height in the visual field than when the jump was only visible as a change in binocular disparity.

Whereas Allison & Howard's (2000b) and Brenner & Smeets' (2006) findings suggest that monocular cues are processed more quickly for estimating slant and distance, Greenwald et al.'s (2005) findings suggest that binocular disparity is processed more quickly for estimating slant changes. The reason for this discrepancy is not clear because the three studies differed considerably in various aspects. For example, Greenwald et al. showed alternating white and black frames for 167 ms before presenting the changed slant, in order to mask the slant change. Allison & Howard and Brenner & Smeets did not mask the perturbations. Furthermore, in Greenwald et al.'s study subjects moved a real object so that the visual information matched the proprioceptive information. In Brenner and Smeets' study the visual position did not match the position that was felt, since subjects moved a cursor with a mouse. Moreover, in Brenner & Smeets' study the perturbation was a change in position, whereas in Greenwald et al.'s it was a change in slant. Such differences make it impossible to tell which aspect is responsible for the different conclusions as to whether binocular information is processed faster or more slowly than monocular information.

In this study, we investigated whether latency differences between responses to changes in binocular disparity and changes in the monocular image are visible in the online control of movement. As in Greenwald et al. (2005), subjects placed a cylinder on a surface of which the slant could change right after movement onset. Either the binocular disparities or the monocular images or both could indicate the change in slant. We determined how subjects altered the orientation of their hand in response to such a slant change. We blanked the screen before the slant changed on half of the trials to determine whether seeing the change allows one to respond faster. On such trials subjects could respond to the new slant but not to the transient.

Methods

Subjects
Eight subjects (three male, five female) participated in the experiment. Five subjects were naïve with respect to the purpose of the experiment. All subjects had normal or corrected-to-normal vision, with a stereo acuity below 1 arcmin (tested with RandotTM plates).

Apparatus
Subjects sat behind a surface (45 cm by 45 cm) that was centered 60 cm in front of the midpoint of their body and 40 cm below their eye-level. This surface could be rotated around a transversal axis with the help of a computer-controlled motor (see Figure 3.1). They held a cylinder with a height of 6 cm and a diameter of 9.5 cm in their right hand. At the subjects' right side, 26 cm from their midsagittal plane, 60 cm in front of them, and 18 cm below eye-level, there was a second surface with a 2mm deep indentation in the shape of the base of the cylinder. Subjects had to place the cylinder within this indentation at the start of each trial.
Subjects did not see the real surface, starting position or cylinder. They saw a virtual surface, starting position and cylinder. The three-dimensional virtual environment was created by presenting different images to the left and right eyes using a combination of two CRTs and mirrors (see Figure 3.1). The mirrors were semi-silvered with occluders attached behind them. We matched the virtual and real environments by removing the occluders and monocularly aligning the corners of a rectangle on the screen (as reflected by the mirror) with the 3D positions of four markers on a calibration rectangle that was placed above the real surface (as seen through the mirror) for that purpose. Using a 3D virtual environment enabled us to dynamically and independently manipulate the virtual surface's slant as specified by binocular disparity and by the monocular images.

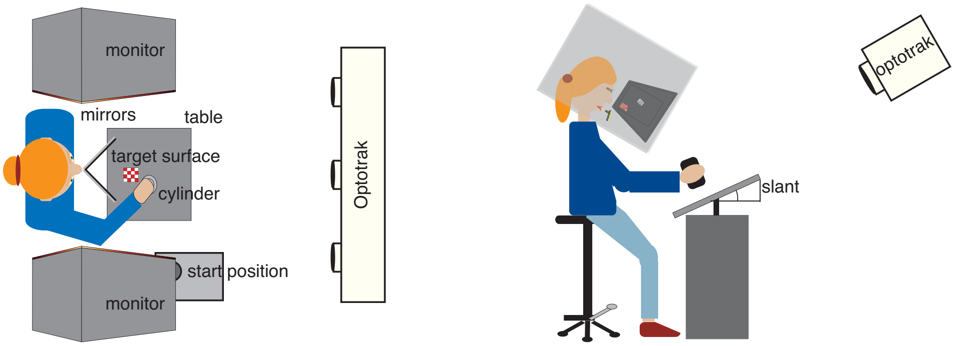

Figure 3.1: Schematic representation of the set-up (not to scale).

Throughout the experiment we recorded the 3D positions of three Infra-red Emitting diodes (IREDs) that were attached to the cylinder using an Optotrak 3020 system (Northern Digital, Inc.), so that we could generate images of the cylinder while the subject was moving it. Since motion parallax as a result of small head movements has been found to contribute to slant perception (Louw et al., 2007), we determined the positions of the eyes relative to a bite-board before the experiment and recorded the position of the bite-board using the Optotrak system during the experiment. The bite-board was not attached to anything, so subjects were free to move their head. We continually adjusted the images to the positions of the subject's eyes so that the slant indicated by motion parallax was consistent with the slant indicated by static information from the monocular images. Note that this refers to the 3D position of the subjects' eyes in space. The direction of gaze was neither monitored nor instructed.

The positions of the IREDs were sampled with a frequency of 250 Hz. Based on the coordinates of the IREDs on the cylinder and bite-board, a PC calculated the current position of the cylinder and eyes and sent these coordinates to two Apple G5's that each rendered an image of the cylinder and virtual surface for one of the eyes on a CRT monitor (1096 * 686 pixels, 47.3 by 30.0 cm). The new images were created with the frequency of the refresh rates of the two CRT monitors (160 Hz). Thus we generated images that were appropriate for the actual position of the eyes and hand at each moment in time. The delay between a cylinder movement and the visual feedback was about 20 ms.

Stimuli

A small pink virtual sphere indicated the starting position. The virtual surface was a square with sides of 10 cm and was visible as a red and grey checkerboard of 4 by 4 squares. The virtual object was a cylinder with 14 white and black stripes and a green top and bottom. It had the same dimensions as the real cylinder. The shapes of the projections of the squares of the checkerboard on the screens provided monocular information about surface slant. The differences between the computer images for the two eyes provided binocular information about surface slant (binocular disparity). Both information sources were also available for the cylinder.

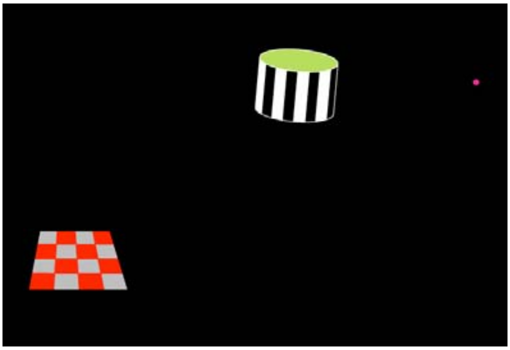

Figure 3.2: Impression of the subject's view during the experiment. From left to right, the virtual surface, the cylinder and the starting point.

Procedure

Each trial began by the computer positioning the real surface in the orientation that the virtual surface would have at the end of the trial. This happened in 3 to 4 movements to prevent subjects from deducing the final orientation on the basis of the sound from the motor that positioned the real surface. Subsequently, the virtual surface and the virtual cylinder were presented. The virtual surface was slanted by 5° from the horizontal plane (positive angles indicate that the side of the virtual surface furthest from the subject's body is higher than the side nearest to the subject's body). Subjects had to move the cylinder to the starting position and, after hearing a beep, to accurately place the cylinder on the virtual surface. The beep was presented 500–800 ms after subjects placed the cylinder at the starting position. We regarded the moment that the cylinder had traveled 20 mm from the start position (in any direction) the moment the subject reacted. If subjects reacted before the beep or within 100 ms after presentation of the beep, the movement was considered to have started too early. The trial was then stopped and presented again.

When subjects reacted, the slant of the virtual surface could change to either -5° or +15°. The binocular disparities could change at the same time as the monocular images (as they normally would) or only one of the two cues could change at that moment. If only one cue changed, the other did so 150 ms later to ensure that both the binocular disparities and the monocular images indicated the same slant at the end of the trial. This final slant was always consistent with the slant of the real surface that subjects felt at contact. When the two cues were in conflict for 150 ms, one cue always still indicated +5° while the other indicated a changed slant (either -5° or +15°).

On half of the trials the surface disappeared at the onset of the beep and only reappeared again at the moment of the first slant change, which meant that no surface was visible for about 620 ms. As a result, subject could not see the change on these trials (i.e. they could not see the transient for the first slant change). Figure 3.3 summarizes the 14 conditions. Each condition was presented at least 25 times to each subject.

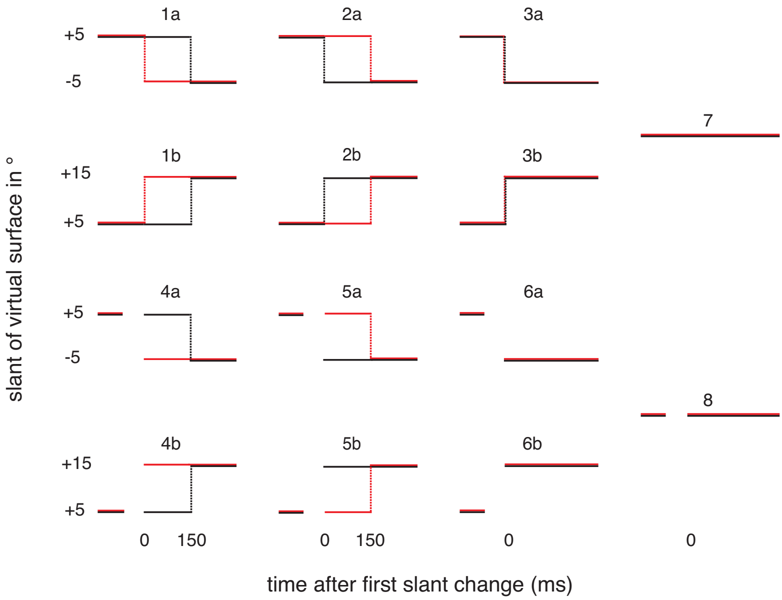

Figure 3.3: The six pairs of conditions in which the slant changed (conditions 1-6) and the two conditions in which it did not (conditions 7 and 8). The virtual surface always had a slant of 5° at the beginning of the trial. Black lines represent the slant indicated by binocular disparity. Red lines represent the slant indicated by the monocular images. Just after movement onset, the slant of the virtual surface could change by -10° (a) or +10° (b). This change was either in the monocular information (pairs 1 and 4), the binocular information (pairs 2 and 5), or both (pairs 3 and 6). If only one changed, the second changed 150 ms later. On half of the trials the image was blanked from the auditory 'go' signal to the first slant change (conditions 4–6 and 8; gap not drawn to scale).

Data Analysis

We used the orientation of the base of the cylinder in relation to the horizontal plane as our measure of cylinder orientation. This orientation was determined from the position data provided by the three IREDs on the cylinder. When one or more of these IREDs were missing, we interpolated their positions from the positions on the frames before and after the frame with the missing markers. We rejected a trial if it had more than 10 frames with missing markers in the period after the first slant change and before making contact with the surface, or if the movement time was more than 1.5 s.

We determined the cylinder's angular velocity by fitting a 2nd order polynomial to the orientations of the cylinder during a 5-frame period centered on each frame, and determining the derivative of this polynomial at that frame (Biegstraaten et al., 2003). We synchronized the trials at the moment of the (first) slant change. For each subject, condition, and frame after the slant change, we then averaged the angular velocities and determined the corresponding standard errors.

Since we found no indication that the sign of the slant change influences the latency of responses to a cue (see Figure 3.4), we analyzed the difference in angular velocity between conditions that have the same timing of the slant changes but a different direction of slant change (e.g. conditions 1a and 1b in Figure 3.3), rather than comparing each with the unperturbed conditions (7 and 8). We determined the onset of the response to the first slant change in three steps. First we searched from the moment of the slant change to find the first frame in which the angular velocities for the two directions of slant change differed by more than 2 standard errors (in this difference). Then we determined the maximal difference in angular velocity during the subsequent 100 ms and searched back from the frame at which this maximum occurred to find the frames at which the difference was 25% and 75% of the maximum. We considered the intersection of the line through these points with a line representing zero velocity difference to be the onset of the response (Veerman et al., 2008). We compared the onsets in the six conditions using a repeated-measures ANOVA and Fisher's PLSD (α=0.05).

Although subjects obviously also responded to the second slant change in conditions in which a second change occurred 150 ms later than the first, we did not analyze such responses. Such an analysis would be quite complicated, because by then subjects are already responding to the first change. The purpose of the second change was only to ensure that the feedback at the end of the trial felt correct. The 150 ms interval was long enough to ensure that we did not inadvertently consider responses to the second change to be late responses to the first.

Results

On average, 56 of the 545 trials per subject were rejected because they contained too many frames with missing markers or because the movement time exceeded 1.5 s. Individual subjects' mean reaction times varied between 490 and 900 ms. Mean movement times, the time from reaction to the moment the cylinder made contact with the surface, ranged from 500 to 900 ms.

Figure 3.4 shows the mean angular velocity of the cylinder over time. Zero on the time (horizontal) axis represents the moment that the slant (first) changed. The data confirm that responses in the two directions are similar (see the symmetry with respect to the unperturbed conditions 7 and 8). There are clear differences between the onsets and shapes of the responses in the different conditions. To get a better view of these differences, we determined the difference between the mean responses in the pairs of conditions that only differ in the sign of the slant change (Figure 3.5). When the transient is present, subjects respond faster to the changes in the monocular images than to the change in binocular disparity. When there is no transient, subjects respond faster to a change in binocular disparity. When the monocular images and the binocular disparity change together, subjects responded as fast as they did to the fastest cue.

When we compare the responses in which the transient was present with responses in which it was not (Figure 3.5), we see that responses to changes in disparity are not affected by removing the transient, but responses to changes in the monocular images are. The responses to changes in the monocular images also appear to initially be weaker when the transient was present, but we cannot be sure of this because more variability in latency could also account for the less abrupt change in the average angular velocity.

Even when responding to the monocular images without the transients (condition 4), subjects responded fast enough to be sure that they were not responding to the change in binocular disparity that occurred 150 ms later. It took subjects about 200 ms to respond to a change in binocular disparity. The responses to a change in the monocular images when the screen had been blanked occurred well before 200 ms after the second change (350 ms after the first change). Thus subjects responded to the changes in the monocular images as well as to the changed monocular images.

These findings show that subjects can use (at least) three types of information to adjust ongoing movements. In order of increasing latency: the changes in the monocular images, the new information about slant from binocular disparity and the new information about slant from the monocular images.

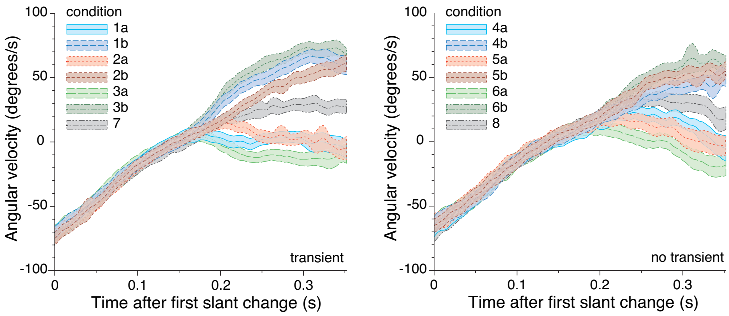

Figure 3.4: Mean velocity of the cylinder's rotation. Time zero represents the moment that the first cue changed (except in conditions 7 and 8 where it is the time that it would have changed). In condition pairs 1 and 4 the monocular images changed first. In condition pairs 2 and 5 binocular disparity changed first. In each pair of conditions the amplitude of the change was either -10º (a) or +10º (b). The shading indicates the standard error between subjects.

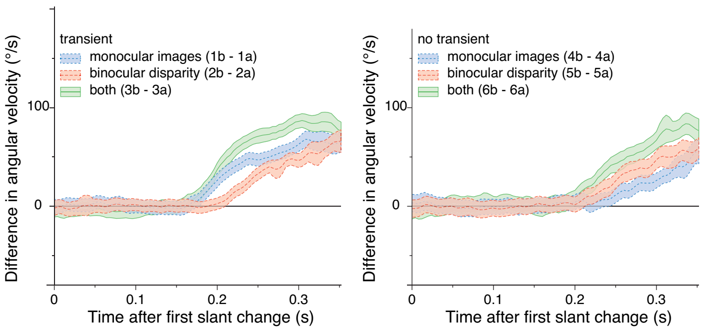

Figure 3.5: Mean difference in angular velocity between conditions that have the same timing but a different sign of the slant change. The shading indicates the standard error between subjects.

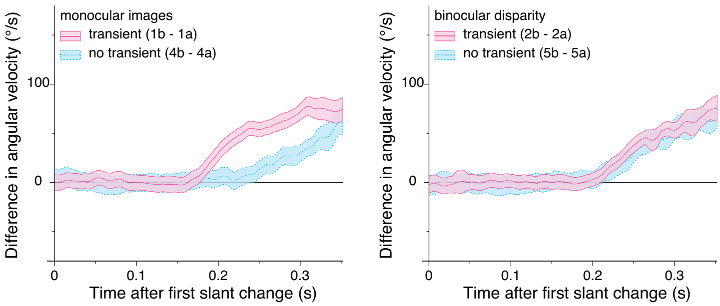

Figure 3.6: Comparing conditions in which the (first) slant change was visible (transient) and conditions in which it was not (no transient). Other details as in Figure 3.5.

To investigate whether the differences that we see in Figures 3.5 and 3.6 are consistent across subjects, we determined the onsets of the responses for each subject (Table 3.1). The averages of these response latencies are shown in Figure 3.7. The data in Table 3.1 and Figure 3.7 suggest that subjects can respond to three sources of information: the changes in the monocular images (condition pairs 1 and 3), the new slant from binocular disparity (2, 5 and 6) and the new slant from the monocular images (4). The ANOVA on the onsets of the responses revealed a significant main effect of Condition ($p < 0.01$). Fisher PLSD tests confirmed the division into the three groups indicated by the rectangles in Figure 3.7. The onsets never differed significantly between members of the same group, and always differed significantly between members of different groups.

Subject \ Pair	1	2	3	4	5	6
1	0.15	0.15	0.14	0.26	0.21	0.21
2	0.17	0.21	0.16	0.20	0.20	0.17
3	0.17	0.20	0.18	0.20	0.22	0.19
4	0.15	0.20	0.20	0.29	0.30	0.19
5	0.19	0.23	0.16	0.30	0.24	0.22
6	0.18	0.21	0.17	0.30	0.16	0.18
7	0.15	0.20	0.17	0.23	0.18	0.19
8	0.17	0.23	0.17	0.27	0.22	0.22

Table 3.1: Estimated latency for each pair of conditions (see Figure 3) for each subject (in seconds).

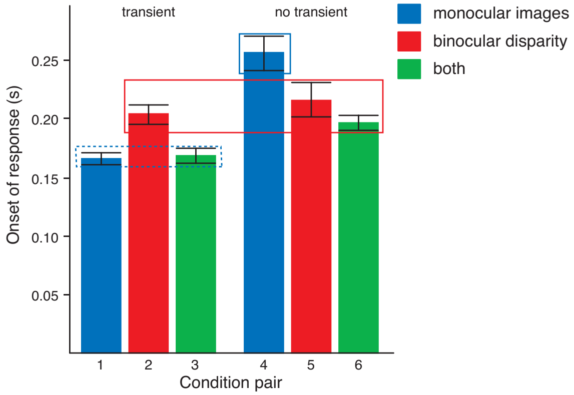

Figure 3.7: Average onset of the response for each pair of conditions (as listed in Figure 3). The blue dotted rectangle indicates responses based on changes in the monocular images. The blue solid rectangle indicates responses based on a new estimate of slant from the monocular images. The red rectangle indicates responses based on a new estimate of slant from binocular disparity.

Discussion

Subjects alter the orientation of their hand to match the changing surface slant during their movement. When an object changes orientation its projection on the retina changes. In our virtual environment, the retinal projections of the squares of the checkerboard surface become more trapezoidal when the far side of the checkerboard moves downwards, because the lateral images sizes and separations decrease for the parts of the surface that move further away and increase for those that move closer. Our data suggests that subjects respond to these changes in each eye's image, as well as to a changed image in each of the eyes.

Response latency depends on saliency (Veerman et al., 2008). However, blanking the screen did not simply increase the latency by reducing the salience of all changes in slant because responses to changes in binocular disparity were not affected by blanking the screen (see figure 6). Thus blanking the screen specifically influenced responses to changes in the monocular image.

Subjects responded 40 ms faster to changes in the monocular images than to new information about slant from binocular disparity, and 90 ms faster to changes in the monocular images than to a new slant from the monocular images. When more than one source of information changed at the same time, the latency of the response was that of the fastest source.

For a change in binocular disparity, the onset and the slope of subjects' responses were almost identical in the conditions with and without a visible transient, indicating that in both conditions subjects responded to the same kind of information. This suggests that subjects do not respond directly to the changes in binocular disparity, but only to the new disparities. This not necessarily mean that such changes are not detected, because they may simply be detected with a longer latency than the disparities themselves.

The latency differences that we find can explain the differences between the studies mentioned in the introduction. The jumps in distance in Brenner & Smeets' study (2006) and the temporal modulation and step changes in Allison and Howard's study (2000b) were clearly visible, so subjects could readily respond to changes in the monocular images. The perturbations in Greenwald et al.'s study (2005) were masked by a set of frames alternating between black and white for 167 ms, so responses were probably based on a changed estimate of slant. We found that responses to a new slant from binocular disparity were faster than responses to a new slant from the monocular images when the transient was not visible, but that responses to changes in the monocular images were even faster (when they were visible).

In our study we intentionally chose natural conditions, such as an approximately horizontal surface in front of and below the subjects’ eyes. We compared changes of the same magnitude for all cues. Under such conditions we find that a change in the monocular images is processed faster than a new binocular slant and that a new binocular slant is processed faster than a new slant from the monocular images. Perhaps conditions can be found in which the order of the latencies is different, but we believe that our findings are representative of many natural circumstances.

Responses to changes in the monocular images appeared not only to be quicker, but also stronger than responses to a new slant (the slope of the pink curve versus the slope of the blue curve in the left plot of Figure 3.6). A possible explanation for this is that people respond to the motion in the retinal image by initiating a strong response in the direction of the change (before knowing when and where the change will end), whereas a new slant only initiates a response that is proportional to the change in slant (and the weight given to the source of information involved). In other words, subjects may initially match the way they rotate their hand to the motion of the surface, and then adjust the end orientation of the ongoing response to the new slant estimates. This combination of responding to a derivative and to a combined slant estimate would nicely integrate optimal accuracy with fast responses. However, this poses a challenge to optimal cue combination theory, because the changes in the monocular images are temporary signals, so they cannot simply be averaged with other information to get a better estimate of slant. Thus, whenever the observer is moving relative to a surface, the whole dynamics of the interaction will have to be considered to predict how slant information is used.

Chapter 4:
Reliability is more than precision: Fast adaptation of slant cues' weights based on their 'correctness'

Abstract

According to optimal cue combination, when two visual cues provide information about the same attribute, their estimates are averaged with weights that are inversely proportional to their variance. As a result, the precision of the combined judgment is maximized. If so, the weights should be independent of whether or not the cues give correct information. To test this, we asked subjects to put a cylinder on a slanted surface. Surface slant was defined by monocular and binocular cues that indicated different slants on most of the trials. Haptic feedback was initially consistent with either the monocular cues or binocular disparity, and then changed to match the other cue. Subjects almost immediately increased the weights of cues that matched the feedback. This quick change in weights indicates that the correctness of the cues is considered, not only their precision.

Introduction

We estimate the orientation of objects relative to our body on the basis of retinal cues such as texture density, image deformation and binocular disparity. Estimates from all of these cues are believed to be combined in a statistically optimal fashion, meaning that the variance of the combined estimate is minimized (Adams & Mamassian, 2004; Ernst & Banks, 2002; Hillis et al., 2004; Jacobs, 1999; Knill, 1998; Knill & Saunders, 2003; Landy et al., 1995; Muller et al., 2007; Oruc et al., 2003; van Beers et al., 1999). Such optimal cue combination consists of weighted averaging, with weights that are inversely proportional to the uncertainty or variance in the value of the cue (Landy et al., 1995; van Beers et al., 1999; Yuille & Bülthoff, 1996). When we refer to optimal cue combination in the remainder of the paper we mean optimal according to this definition.

Knowledge about the uncertainty associated with a cue could be based on prior experience. However, because the variance in the estimate of a cue depends on many factors (e.g. slant, distance, texture) all these factors would have to be considered. A simpler method would be to estimate the cue's precision from the precision of the coding. Less precise information is likely to activate a broader range of cells; i.e. to stimulate cells that are sensitive to a wider range of possible values of the property that is estimated. The precision of a cue's estimate can therefore be determined from the spread of activity within neuronal populations that are sensitive to that cue. Irrespective of whether the weight given to a cue when combined with other cues is based on such instantaneous neuronal tuning or on prior experience, whether or not the estimate is veridical (correct) is irrelevant.

Several studies indicate that the cues' variances do not fully determine their weights. Ernst et al. (2000) found that subjects decreased the weight that they gave to either binocular disparity or texture when making slant judgments based on a combination of the two if the value of that cue did not match the felt slant of a surface during a training phase in which they moved a cube up and down across the surface. Similar results have been obtained for combinations of cues for size and shape (Atkins et al., 2001; Jacobs & Fine, 1999). These studies indicate that subjects can learn that one cue is less accurate than others from haptic or auditory feedback, and can use this information to adapt the weights given to the cues accordingly. Since in all these cases the precision of the cues remained constant, such reweighting seems to be inconsistent with the notion of optimal combination only depending on precision.

However, Knill (2007) found that cues' weights could change in a similar manner without haptic or auditory feedback. In his study there was also no change in the physical precision of the visual information presented to the subjects. People need to make assumptions in order to use certain cues. One such assumption is that elliptical images on our retinas are projections of slanted circles in the real world. Knill proposed that people change their assumptions to conform to experience within a given context, so that the judged precision of a cue's estimate, and hence its weight, is influenced by experience that is inconsistent with that in daily life. In Knill's experiment subjects were exposed to many non-circular elliptical shapes. They later relied less on foreshortening and more on binocular disparity when tested in a similar context. The change in weights developed gradually (over days) with experience. Knill argued that being less biased towards perceiving the shapes as being circular made subjects consider a broader range of slants for a given retinal shape, resulting in the foreshortening cue having a lower precision and therefore being given less weight.

In the present study we investigated the time course of the changes in the weight *given* to monocular cues and binocular disparity when haptic feedback indicates that one of them is biased. We were interested in the time course because this might give insight into the mechanism by which the weights change. In the previous studies in which weights given to complementary cues were influenced by haptic or auditory feedback (Atkins et al., 2001; Ernst et al., 2000; Jacobs & Fine, 1999), the authors assumed that many trials were needed to change the weights, and the experiments were designed accordingly. Since in daily life inconsistencies between complementary cues are presumably primarily due to random variability in the individual estimates, it makes sense for the changes in the weights to be slow, because there is no point adjusting the weights to random variability. Moreover to be able to quickly change the weights in accordance with the feedback, one must also have access to the individual cues' estimates, not only access to the combined estimate. A change in weights based on information about likely shapes within a certain context (Knill, 2007) is also necessarily slow. We therefore decided to examine whether the changes in weights in response to the feedback are really slow, and if so how slow.

Subjects sat in a 3D virtual environment in which we could independently manipulate surface slant from monocular and binocular cues. They had to put a cylinder on the center of a virtual surface. Monocular and binocular cues either indicated the same slant or slants that differed by 20º. A real table was rotated to match the slant of one of the cues, so that at the end of their movement the subjects received feedback from the felt slant of the table that was consistent with the slant indicated by one of the cues. Previous studies were divided into separate learning and testing stages (Atkins et al., 2001; Ernst et al., 2000; Jacobs & Fine, 1999; Knill, 2007). Because we were interested in the time course of the change in weights, we devised a paradigm in which we could examine how the weights changed within a block of learning trials instead of using separate test and learning stages. We introduced a switch in feedback during a session. Subjects first performed a short block in which the table had the same slant as one of the cues, and then a long block in which the table had the same slant as the other cue. We calculated the weight of the monocular cue over groups of trials within each block to determine the time course and the direction of the change in weight.

Method

Subjects

Nine experienced subjects (four male, five female) participated in the experiment, including two of the authors. All subjects had normal or corrected-to-normal vision, with a stereo acuity better than 60 arc seconds (tested with Randot StereoFlyTM).

Apparatus

Subjects sat behind a surface (45 cm by 45 cm), centered 60 cm in front of their chest and 40 cm below their eye-level. They held a cylinder with a height of 6 cm and a diameter of 9.5 cm in their right hand. At the subjects' right side, 26 cm from their sagittal plane, 60 cm in front of them and 18 cm below eye-level, there was a second surface with a 2mm deep indentation in the shape of the base of the cylinder. Subjects had to place the cylinder within this indentation at the start of each trial. Subjects did not see the real surface, starting position or cylinder, but only a virtual surface, starting position and cylinder.

The three-dimensional (3D) virtual environment was created by presenting different images to the left and right eyes using a combination of two CRTs and mirrors (see Figure 4.1). The mirrors were semi-silvered with two occluders attached behind them. We matched the virtual and real environments by removing the occluders and monocularly aligning the corners of a rectangle on the screen (as reflected by the mirror) with the positions of four markers on a calibration rectangle that was above the table (as seen through the mirror). By using a virtual environment we could dynamically manipulate the slant of the virtual surface independently for monocular and binocular cues (as explained below).

We recorded the 3D positions of three Infra-red Emitting diodes (IREDs) that were attached to the cylinder with an Optotrak 3020 system (Northern Digital, Inc.). The positions were recorded throughout the experiment, so that we could generate images of the cylinder while the subject was moving it. Motion parallax as a result of small head movements has been found to contribute to slant perception (Louw et al., 2007). We therefore determined the positions of the subjects' eyes throughout the experiment so that the appropriate motion parallax was present in our stimuli. To enable us to do so, subjects were fitted with a bite-board with 3 IREDs. The bite-board did not restrain the subjects because it was only attached to the head. We determined the positions of the eyes relative to the bite-board before the experiment started and recorded the bite-board's position during the experiment. In this manner, we could continually adjust the images to the position of the subject's eyes without restraining the head.

Based on the coordinates of the IREDs on the cylinder and bite-board (sampled at a frequency of 250 Hz), a PC calculated the current position of the cylinder and eyes and sent these coordinates to two Apple G5's that each rendered an image of the cylinder and virtual surface for one of the eyes on a CRT monitor (1096 x 686 pixels, 47.3 x 30.0 cm). The new images were created with the frequency of the refresh rates of the two CRT monitors (160 Hz). Thus we generated images that were appropriate for the actual position of the eyes and hand at each moment in time. The delay between a cylinder movement and the visual feedback was about 20 ms.

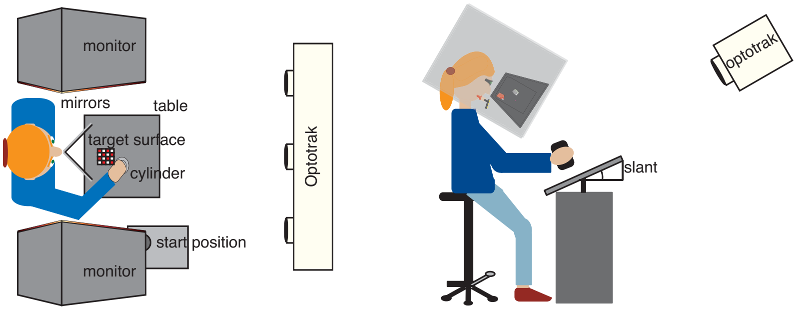

Figure 4.1: Schematic representation of the set-up (not to scale)

Stimuli

A small pink virtual sphere indicated the starting position. The virtual surface was a square with sides of 10 cm. It had a red and grey checkerboard pattern of 4 by 4 squares. The virtual cylinder had 14 white and black stripes along its main axis and a green top and bottom. It had the same dimensions as the real cylinder. The shapes of the projections of the squares of the checkerboard on the screens provided monocular information about surface slant. Motion parallax was always consistent with this slant. The differences between the computer images for the two eyes provided binocular information about surface slant (binocular disparity). All sources of information were also available for the cylinder.

Each cue specified a slant of either -5° or +15° for the virtual surface. Positive values indicate that the side of the checkerboard that is farthest from the subject's body is higher than the nearest side. The monocular and binocular cues either indicated the same slant (non-conflict trials) or different slants (conflict trials). In order to render different slants for the two cues, we determined how a square surface with a slant defined by the monocular cues would look for a single (cyclopean) eye, and then rendered images for the two eyes that (on average) provide this retinal image, while having the binocular slant that we wanted (Knill, 1998; Landy et al., 1995; Van Mierlo et al., 2007). This is equivalent to a simulation of a deformable non-conflict trapezoid with a slant as specified by binocular disparity.

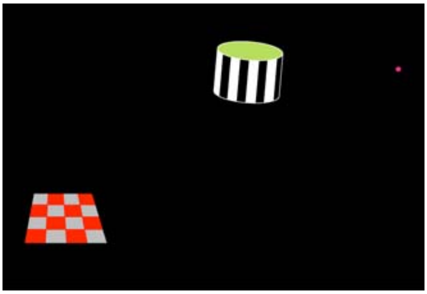

Figure 4.2: Impression of the subject's view during the experiment, with from left to right the virtual surface, cylinder and starting point.

Procedure

Before each trial the computer positioned the table at an orientation of either +5° or -15° from the horizontal plane. The positioning happened in 3 to 4 movements to prevent subjects from deducing the slant of the table from the sound that the motor made when positioning the table in the correct orientation. Subsequently, the virtual surface, the starting point and the virtual cylinder were presented.

Subjects started a trial by positioning the cylinder at the starting position. A beep was presented 500 – 800 ms after they did so. This was the signal that they should move the cylinder to the center of the virtual checkerboard. They were asked to do so as accurately as possible in one single continuous movement. If subjects responded before the beep, or within 100 ms after presentation of the beep, the movement was considered to have started too early. The trial was then stopped and the subject had to start again. The table was rotated to match the slant of one or both of the cues, so it gave feedback about the correctness of the end orientation of the subjects' hand and thus about the accuracy of the visual information that was used to make the movement.

In Ernst et al. (2000) the weights given to texture and disparity differed more between the two pretest blocks than between the pretest and posttest of a conflict block. To minimize such idiosyncratic behavior, we introduced the change in the feedback during the session itself. The change took place after 30 trials. We studied how subjects changed the weights in the remaining 102 trials. We repeated the experiment twice. In session A, the felt slant matched that of binocular disparity in the first 30 trials, and that of the monocular cues in the 102 trials that followed. In session B, the felt slant matched that of the monocular cues in the first 30 trials and that of binocular disparity in the remaining 102 trials. The order of the two sessions was counterbalanced across subjects and the sessions were performed on different days. For each experiment, the first 30 trials contained 20 conflict and 10 non-conflict trials. The remaining 102 trials contained 68 conflict and 34 non-conflict trials. The four combinations of slants (two conflict and two non-conflict combinations) were presented in random order.

Data Analysis

Our main variable was how much the slant indicated by each cue contributed to the perceived slant of the virtual surface. As a measure for perceived slant we took the orientation of the cylinder just before it made contact with the table, which we will refer to as its end orientation.

We determined the cylinder's orientation in the direction of the perceived slant of the virtual surface: the slant of the base of the cylinder relative to the horizontal plane. This slant was determined for each frame of each trial from the position data of the three IREDs on the cylinder. If data from one of the IREDs was missing, we interpolated its position from those on the previous and next frames. We rejected a trial if it had more than 10 frames with missing markers.

As the end orientation of the cylinder we took its slant when it was 1.5 cm from where it stood still on the checkerboard surface. This moment was determined on the basis of the position of one of the three IREDs on the cylinder. We searched for a 100-frame period during which the position of this IRED varied less than 0.075 mm between frames. We then determined on which frame the IRED was 1.5 cm from the position at standstill. The orientation on this frame was taken as the end orientation of the cylinder. Figure 4.3 shows an example of this procedure for one trial.

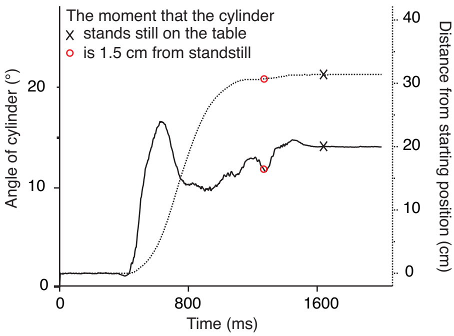

Figure 4.3: Data of one trial illustrating the procedure that we used to determine the end orientation of the cylinder. In this trial both cues indicated a +15 degree slant. The continuous line shows the angle of the cylinder. The dotted line shows the distance from the starting position. The black crosses show the distance and orientation of the cylinder at the moment it came to a standstill on the table. The red circles show these values when the cylinder was 1.5 cm from this position. The red circle on the continuous line indicates the end orientation.

As a compromise between the temporal resolution that we wanted to achieve and the need to average across trials to remove unsystematic variability in the movement (see continuous line in Figure 4.3), we divided each session into ten groups of trials and averaged the trials of the same condition within each group of trials. The short block was divided into 2 groups. The long block was divided into 8 groups. We also calculated the standard errors of the mean end orientation for each group of trials. After rejecting trials with too many missing markers (see above), at least 8 conflict trials and 4 non-conflict trials remained in each group.

We determined the weight of the monocular cues (w_m) in determining the end orientation of the cylinder for each subject and each group of trials with the help of the equation for cue combination by weighted averaging:

$$s_c = w_m s_m + (1 - w_m) s_b \tag{4.1}$$

which can be rewritten as:

$$w_m = \frac{s_c - s_b}{s_m - s_b} \tag{4.2}$$

In Equations 4.1 and 4.2, s_c is the end orientation of the cylinder in one of the conflict conditions. The slants indicated by the individual cues (s_m and s_b) are mean end orientations of the cylinder on non-conflict trials. S_m is the mean end orientation of the cylinder in the non-conflict trials in which the two cues indicated the same slant as the monocular cues in the conflict condition. S_b is the mean end orientation of the cylinder in the non-conflict trials in which the two cues indicated the same slant as binocular disparity in the conflict condition. By using s_m and s_b rather than the true slants we account for issues such as the fact that we determine the angle some time before the cylinder reaches the surface.

We also calculated the standard error for the weight given to the monocular cues (SE_{w_m}) using Equation 4.3:

$$\left(\frac{SE_{w_m}}{w_m}\right)^2 = \left(\frac{\sqrt{SE_c{}^2 + SE_b{}^2}}{s_c - s_b}\right)^2 + \left(\frac{\sqrt{SE_m{}^2 + SE_b{}^2}}{s_m - s_b}\right)^2 \tag{4.3}$$

in which SE_c, SE_b and SE_m represent the standard errors in the settings in the same trials as for Equations 4.1 and 4.2.

For each subject and each group of trials, we averaged the values of w_m and SE_{w_m} across the two types of conflict trials (binocular disparity indicating a slant of +15° and the monocular cues -5°; binocular disparity indicating a slant of -5° and the monocular cues +15°).

Since we found considerable differences between the calculated values of SE_{w_m} for different subjects, we averaged w_m across subjects using weights that were proportional to the subject's precision. For each group of trials, we multiplied each subject's values of w_m by the value of i_w described by Equation 4.4 (where ${}^{i}SE_{w_m}$ is the value of SE_{w_m} for that individual subject) and then summed these values across subjects. We determined the combined SE_{w_m} using the same weights.

$$i_w = \frac{1}{{}^{i}SE_{w_m}{}^{2} \sum_{i=1}^{9} \frac{1}{{}^{i}SE_{w_m}{}^{2}}} \qquad (4.4)$$

In the calculations above we averaged the non-conflict conditions across the entire experiment to obtain s_b and s_m. To check whether this was valid we examined the end orientations of the different non-conflict conditions for the ten groups of trials. If subjects respond to the feedback by relying less on the current visual information, the end orientation seen on the non-conflict conditions will change during the course of the experiment. They may for instance start relying more on the slant felt on previous trials when they notice that their percepts are inaccurate, especially if they cannot determine why this is the case. We determined the end orientations in the non-conflict conditions for each subject and each group of trials. We then summed these values across subjects with weights that reflect subjects' precision in the non-conflict condition.

Results

Figure 4.4 shows the end orientations of the cylinder in the different non-conflict conditions. The slant was stable throughout the experiment, suggesting that over trials subjects relied on the visual information in the same way.

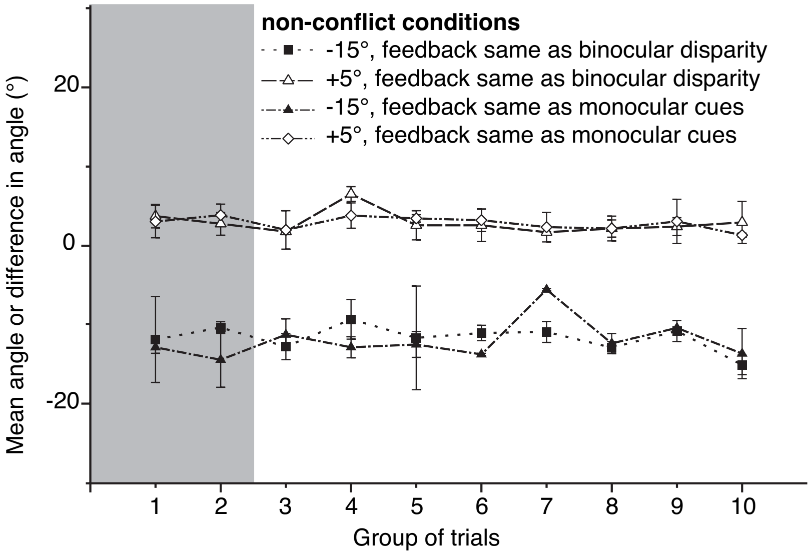

Figure 4.4: The mean slant seen just before the cylinder made contact with the table on the different non-conflict conditions in which the monocular cues and binocular disparity indicated the same slant. The grey region represents the short block of each session. In this block, the feedback differed from that in the remaining trials. In this and all subsequent graphs the error bars show standard errors of the mean.

Figure 4.5 shows the mean weight that subjects gave to the monocular cues for each group of trials. The solid line represents the weight given to the monocular cues in the session in which the haptic feedback initially matched the slant of binocular disparity (open circles on grey background) and then matched the slant of the monocular cues (solid circles on white background). The dashed line represents the same weight in the session in which the feedback first matched the slant of the monocular cues (solid circles on grey background) and then matched the slant of binocular disparity (white circles on white background). After the change in feedback, the weight that subjects gave to the monocular cues changed in accordance with the feedback. It did so within a few trials (remember that each group of trials only contains about 8 conflict trials). Subjects increased the weight given to the monocular cues when it matched the feedback and decreased it when it did not.

To determine whether this pattern of findings is consistent across subjects, we determined the weight that each subject gave to monocular cues during the last 6 groups of trials of each session. We plotted the difference between the weights for the two kinds of feedback (averaged across groups 5-10) in Figure 4.6a. The differences in weight are in the direction that is supported by the feedback for all subjects.

For trials in which the monocular and binocular cues indicated conflicting slants, the simulated shape of the virtual surface was a slanted trapezoid. For trials in which the two cues had the same slant the simulated shape was a slanted square. To see whether subjects quickly adjusted the weights given to the slant cues to the shape seen on the previous trial (similarly to the slow changes found by Knill, 2007), we determined the average weight that subjects gave the monocular cues in the last six groups of trials of the two feedback sessions for conflict trials that were preceded by either a non-conflict trial (black bars in Figure 4.6b) or a conflict trial (white bars). If subjects had more confidence that the surface was square after seeing a square surface (in a non-conflict trial) than after seeing a non-square surface (in a conflict trial), subjects would have decreased the weight of the monocular cues after each conflict trial and increased its weight after each non-conflict trial, so the black bars would be higher than the white bars in Figure 4.6b. This is clearly not the case.

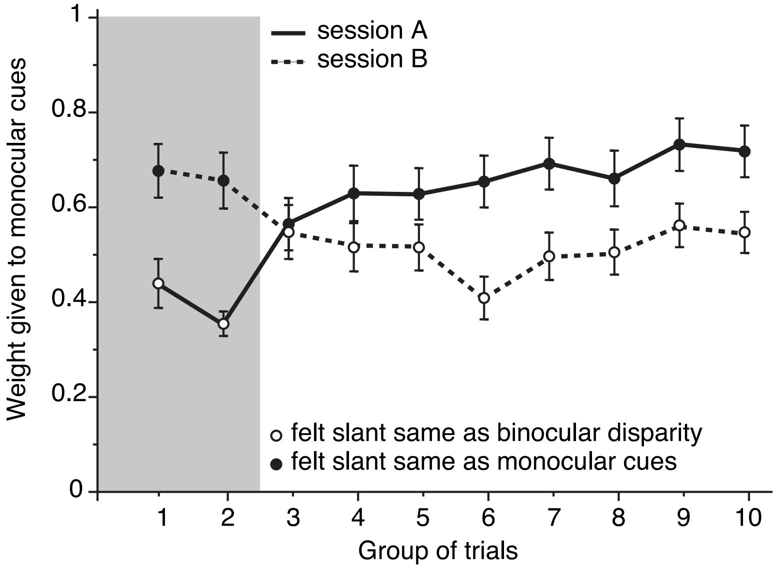

Figure 4.5: Mean weight given to monocular cues in each group of trials. Each group consists of at least 8 conflict and 4 non-conflict trials.

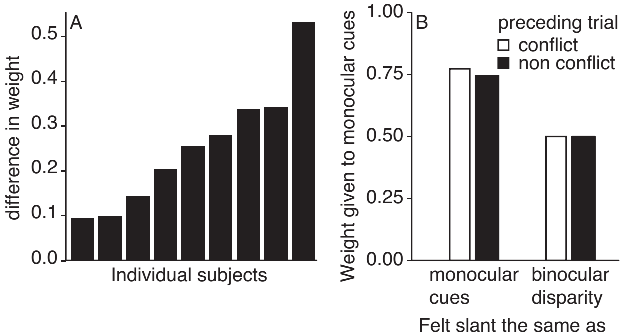

Figure 4.6: A. The difference between the weight given to monocular cues when the feedback was consistent with the monocular cues and when it was not. All subjects adjusted the weights in accordance with the feedback. B. Mean weight given to monocular cues on conflict trials that were preceded by either non-conflict trials (black bars) or conflict trials (white bars), averaged across subjects.

Discussion

Subjects changed the weight that they gave to monocular cues in accordance with the slant that they felt from the table at the end of their movement. If the slant of the monocular cues matched the felt slant in previous trials, they gave the monocular cues more weight than if the slant of the monocular cues did not match the felt slant. We thus reproduced the findings of Ernst et al. (2000).

Ernst et al. (2000), Knill (2007), Jacobs & Fine (1999) and Atkins et al. (2001) all exposed their subjects to many learning trials before judging whether the weights had changed, because they expected the weights to change slowly. In this study we show that the weights given to slant cues can change within just a few trials. The time constant of the change in weights is in line with that found by Smeets et al. (2006) for weights of proprioceptive and visual information after withdrawal of visual feedback about the position of the hand.

Such a fast change in weights is unlikely to result from changes in subjects' expectations about the shapes of objects in the world, or even within the current context. Moreover, if subjects were quickly changing their expectations about whether or not objects were likely to have been squares, we would expect large differences between the weights just after a conflict (non-square shape) or non-conflict (square shape) trial. We found no difference in weight between trials that were preceded by non-conflict trials or conflict trials. Considering how quickly subjects responded to the change in haptic feedback, it is also evident that no elaborate learning mechanisms are involved. We can therefore conclude that a change in judged precision alone cannot explain the change in weights.

Furthermore, the non-conflict data that is presented in Figure 4.4 shows that it is unlikely that the change in weights was brought about by a recalibration of the cues. If subjects changed their interpretation of the values of the individual cues after the change in feedback (as when rescaling the cue) there is no reason to expect them not to do this in the non-conflict trials. So in case of a recalibration, a gradual change in the end orientation should also be seen on the non-conflict trials. No such change was seen, indicating that subjects' interpretation of the cues remained constant throughout the experiment.

Finally, Figure 4.4 shows that the change in weights is not caused by a change in response bias and Figure 4.7 shows that although there were only two possible end orientations of the table, the distributions of responses are unimodal, so subjects were not simply switching between the slants that they had previously felt.

So subjects do not only consider the cues' precision when assigning weights, but also whether the values are likely to be correct. This is in conflict with many current theories that state that the brain integrates different sources of information about the same attribute in a statistically optimal manner (as defined in the introduction). Even assuming that optimal cue combination only takes place when the discrepancies between the cues are not very large (Landy et al., 1995; Muller et al., 2007, 2008; van Ee et al., 2003) does not explain the change in weights in response to feedback about the correctness of the cues.

The extremely fast changes in weights that we found suggests that the values of the individual cues are not lost by combining them into one estimate. Access to the information from the individual cues would be very convenient if subjects are to quickly discover systematic biases and change the weights accordingly, because if subjects would only have access to the combined percept they would not know which cue's estimate was 'wrong'. Thus apparently we do not only have access to an optimally weighted average of the cues. Although optimal weighted averaging describes performance in many experiments extremely well, our findings can only be explained if people can judge whether or not the values of individual cues are correct and consider this, next to the value's precision, when combining different cues into one estimate.

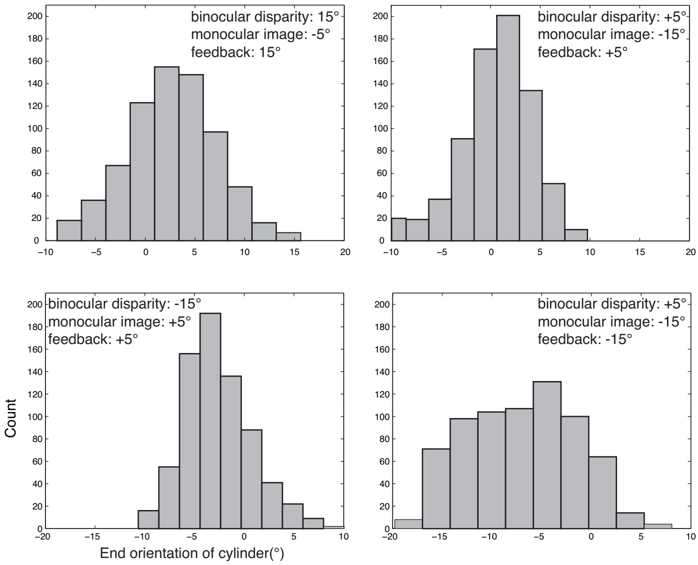

Figure 4.7: Histograms of the cylinder's end orientation for the four conflict conditions.

Chapter 5:
Motion in depth from interocular differences in relative direction of motion

Abstract

When studying manual responses to changes in slant we found that a stereoblind subject responded to changes in binocular disparity. To find out how this was possible we asked three stereoblind subjects to judge the direction in which a horizontal transparent virtual cylinder was rotating. The stereoblind subjects performed better binocularly than monocularly. In contrast to the controls, the stereoblind subjects were also better when the dots defining the cylinder were different for the two eyes. Next, we examined the influence of various perturbations on how they judged the direction of rotation from motion differences between the eyes. Although all perturbations influenced the subjects' performance, performance only dropped to the monocular level when the cylinder expanded and contracted along its axis as it rotated. We conclude that stereoblind subjects use interocular differences in relative direction of motion to judge motion in depth, and do so without matching individual points in the two eyes.

Introduction

If small children display a tendency towards amblyopia, their good eye is usually patched to stimulate the development of the other eye. Long periods of patching, however, may result in abnormal development of binocular vision (McKee et al., 2003). Indeed, one of the authors of this study (JS) received such eye patch treatment during childhood. He shows no detectable binocular vision when tested with the Randot StereoFlyTM or with various kinds of random dot stereograms. JS participated in many previous experiments about binocular vision to check for monocular artifacts. There, JS was unable to align two lines in depth (Brenner & Smeets, 2000), relied completely on the monocular cues when matching the slant of a surface for which monocular and binocular cues indicated different slants (Muller et al., 2007), and failed to adjust his ongoing movement when a target jumped binocularly in depth (Brenner & Smeets, 2006). These findings confirmed that JS is completely unable to use binocular disparities to judge distance or slant. By contrast, in a recent study in which we investigated subjects' online correction of ongoing movement in response to changes in surface slant (Van Mierlo et al., submitted), JS did respond to changes in binocular disparity (Figure 5.1). Why could he do that?

When (part of) an object is moving towards you, you can tell that this is the case from the changing vergence that is required to keep your eyes on it, from its changing disparity relative to other static objects and of course from monocular cues such as changing image size (expansion). There is also some evidence that subjects with normal binocular vision use interocular velocity differences (differences between motion of the object's images on the two retinas) to determine the direction and speed of motion in depth (Brooks & Stone, 2006; but see Cumming & Parker, 1994; Rokers *et al.*, 2008). A special characteristic of this binocular cue is that it does not necessarily require one to precisely identify matching points in the two eyes. The use of a cue that does not require point-by-point correspondence between the eyes is supported by findings that some strabismus patients with low stereo-acuity could also determine the direction of motion in depth in dynamic displays findings (Kitaoji & Toyama, 1987; Maeda et al., 1999).

In the present study, we show that JS and two other stereoblind subjects use such interocular velocity differences to judge the direction of motion in depth. We specifically argue that they use interocular differences in the relative direction of motion to do so.

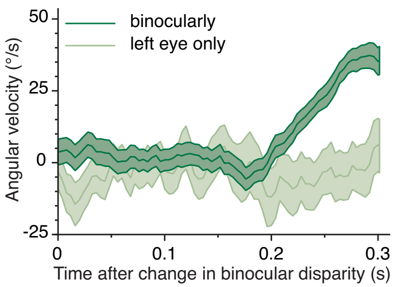

Figure 5.1: JS' response to change in the binocular disparity defined slant of a surface on which he was placing an object. Average velocity of rotation of the object that is being placed (with the standard error across trials). Positive values indicate rotation in the direction of the change in surface slant.

Experiment 1

Previous studies have suggested that estimates of the speed of motion in depth combine motion signals with estimates of displacement (Brenner et al., 1996). To circumvent any influence of displacement, we presented subjects with a transparent virtual cylinder that was defined by randomly distributed limited lifetime dots. The cylinder rotated around its horizontal axis (a screenshot of the stimulus is provided in Figure 5.2). Subjects had to indicate whether dots on the near side were moving upwards or downwards. In order to avoid cue conflicts we used veridical perspective projection. As a consequence, the direction of the rotation could not only be estimated from changing disparity and interocular velocity differences, but also from monocular motion signals. To dissociate between these three cues, we compared performance when the dots were correlated in the two eyes (as they normally are), with performance when the dots had different random positions for one eye than for the other, and with performance when the cylinder was only presented to one eye. Removing the correlation between dots in the two eyes interferes with changing disparity as a cue to motion in depth and removes the static depth information from relative disparities, but it does not necessarily disrupt the use of interocular velocity differences.

Methods

Subjects

We tested three stereoblind subjects (stereo-acuity below the range of the StereoFlyTM test) and seven subjects with normal binocular vision (stereo-acuity better than 80 arcseconds). All three stereoblind subjects had been patched during childhood. The treatment was successful in as far as they reported normal acuity for both eyes.

Apparatus

The stimuli were presented on a CRT monitor (1096 × 686 pixels, 47.3 × 30.0 cm). Subjects sat 70 cm from the monitor, wearing liquid crystal shutter spectacles that successively blocked each eye in synchrony with the refresh rate of the monitor (160 Hz), so that different images could be shown to the left and right eye in rapid alternation. A new image was presented to each eye every 12.5 ms (80 Hz). The subject's interocular distance was taken into account when creating the images. We simulated the cylinder at screen distance (70 cm), so that not only the ocular convergence required to fixate the cylinder and the retinal images were appropriate, but also the accommodation.

Stimuli

The cylinder was defined by 275 red limited lifetime dots. It had a diameter of 8 cm (6.5°) and extended horizontally across the entire screen (24°). The dots always had a diameter of 4 pixels, so that the near and the far side of the cylinder could not be recognized on the basis of dot size. On each trial the cylinder rotated at 20°/s for 2 s and then disappeared. During the 2 s, the dots were asynchronously replaced every 250 ms. New dot positions were chosen at random on the cylinder's surface.

There were four conditions in experiment 1: static, correlated, monocular, and uncorrelated. In the static condition, the cylinder did not rotate. The simulated positions of the dots defining the cylinder were the same for the two eyes. We placed a target dot with twice the diameter of the normal dots (8 pixels) at a random position on a horizontal line through the most distant or nearest part of the cylinder (within 10 cm of the screen center). In the other three conditions the cylinder did rotate. For the correlated condition, the simulated position of each dot was the same for both eyes. For the monocular condition, the image for one of the eyes was black. For the uncorrelated condition, we used twice as many dots, but showed each dot to only one eye.

Procedure

For the correlated, uncorrelated and monocular conditions, subjects had to indicate in which direction the dots were moving on the near half of the cylinder. They pressed the upward arrow key if they thought that the dots moved upwards, and the downward arrow key if they thought that the dots moved downward. No feedback was given, to avoid that subjects might learn to use one of the cues better, which might result them either ignoring or downsizing the importance of the other cues. The conditions were tested in separate blocks of trials. Within each block, subjects were presented hundred trials per direction of movement, in random order. In the monocular condition subjects used their preferred eye.

For the static condition, subjects had to indicate whether the larger dot was on the near or far side. If they thought that the target dot was located on the front of the cylinder, they pressed the downward arrow key. If they thought that it was located on the back of the cylinder they pressed the upward arrow key. There were hundred trials in which the dot was located on the front of the cylinder, and hundred trials in which the dot was located at the back of the cylinder. No feedback was given.

Figure 5.2: One frame (monocular) of the transparent virtual cylinder.

Results

In the static cylinder condition, all three stereoblind subjects performed at chance level, confirming that they could not use disparity to judge the depth of the dots (Figure 2.3). All control subjects performed at almost 100% correct in this condition, as we would expect considering the disparity.

There was a lot of variability in how subjects performed in the monocular condition. This was both so for the stereoblind subjects and for the controls. Some subjects performed well above chance level whereas others did not. On average, both stereoblind and control subjects scored higher than chance level in the monocular condition (Figure 2.4).

The stereoblind subjects performed much better in the uncorrelated condition than in the monocular condition. The control subjects performed at the same level in the two conditions.

In the correlated condition, the control subjects' performance was almost perfect (much better than in the uncorrelated condition). For the stereoblind subjects, the correlation between the two eyes' images made no difference.

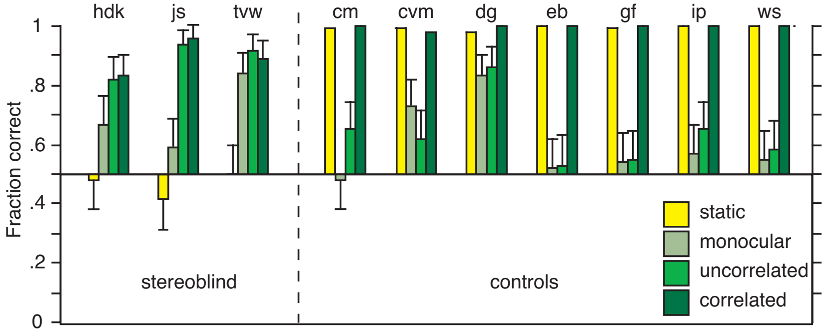

Figure 2.3: Individual subjects' performance. Error bars represent 95% confidence intervals of the mean.

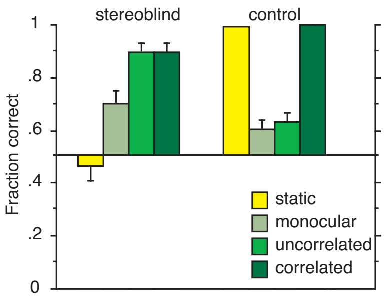

Figure 2.4: Average performance of the 3 stereoblind subjects and of the 7 control subjects. Error bars represent the 95% confidence intervals (between subjects).

Discussion

We can conclude from these findings that the stereoblind subjects use some kind of binocular information that does not require matching points in the two eyes to discriminate between the directions of rotation.

Several subjects performed well above chance level in the monocular condition, indicating that they were able to use the monocular information in our stimulus. A disadvantage of good performance in the monocular condition is that improvements from binocular information cannot be very large. This is the case for the data of stereoblind subject TvW and control subject DG. Despite being able to perform well on the basis of monocular information, so that having binocular information makes little difference, the overall pattern in their data is similar to that of the other subjects.

Note that although completely ambiguous monocular stimuli can be made (using orthogonal projection), this is incompatible with providing exactly the same information to the eye in the other conditions. An advantage of using correct perspective projection is that we need not worry as much about cue conflicts. However, having uncorrelated positions of the dots in the two eyes' images introduces a conflict between the shape that is defined by the random disparities and the (cylindrical) shape that is defined by the motion cues. That the control subjects did not perform better in the uncorrelated condition than in the monocular condition may seem to indicate that they are less adept at using the difference in motion between the eyes than the stereoblind subjects. However, it might also be due to an inability to ignore the random disparities that distort their shape perception. Thus the stereoblind subjects may simply perform better in the uncorrelated condition because they have no conflict with disparity. Similar conflicts may explain why Cumming & Parker (1994) found no evidence that people could use interocular velocity differences to judge motion in depth. In their study interocular velocity differences were in conflict with monocular cues to motion in depth (in particular expansion).

The stereoblind subjects' binocular performance is not affected by whether or not the images in the two eyes are correlated. The fact that our cylinders were transparent implies that subjects cannot determine the direction of the rotation by averaging velocities in particular parts of the visual field and then comparing these averages across the two eyes. The motion of the dots on the front and back plane would then cancel each other. Thus the dots must be grouped before their motion in the two eyes can be compared. The control subjects can do this on the basis of depth information from disparity or on the basis of whether the points were moving upwards or downwards (as suggested by Brooks & Stone, 2006). The stereoblind subjects only have the latter option.

In order to get an idea of what exact attribute of the dots' motion the stereoblind subjects use to perform our task, we conducted a second experiment in which we perturbed various aspects of the motion.

Experiment 2

It is unclear what information the stereoblind subjects used to discriminate between the directions of rotation of the cylinder. They might have been comparing the dots' directions, speeds or accelerations within corresponding parts of the retinas. We perturbed different aspects of the motion and studied how this affected the stereoblind subjects' performance. In detail, we disrupted the dots' paths by translating the cylinder as a whole along its horizontal axis as it rotated around that axis. Doing so makes it impossible to use local motion to determine the direction of the cylinder's rotation. Secondly, we perturbed the dots' relative paths by stretching and compressing the rotating cylinder along its horizontal axis. This not only makes it impossible to use local motion to determine the direction of the rotation, but it also affects the direction of motion of each dot differently, so that the dots' relative motion changes. Finally, we let the cylinder rotate at variable rates that changed asynchronously in the two eyes. This disrupts the pattern of acceleration of all dots within each eye as well as any comparison of velocities between the eyes.

The translation and expansion could either be in the same direction in the two eyes (synchronous conditions), or in opposite directions (asynchronous conditions). The former leaves the relative motion in the two eyes (but not necessarily that between the points in each eye) intact, whereas the latter disrupts the relationship between the motion (asynchronous translation) or even the relative motion (asynchronous expansion) in the two eyes.

For all perturbations, we also examined the influence on monocular performance. The images for the two eyes were always uncorrelated. We only tested the stereoblind subjects.

Methods

We used the same virtual cylinder as in Experiment 1. For the translation conditions, the rotating cylinder oscillated sinusoidally (140°/s) along its horizontal axis with a peak amplitude of 0.5 cm. For the expansion conditions the cylinder expanded and contracted to 110% and 90% of its original size at the same rate. For the synchronous conditions, the translation or expansion was in phase for the two eyes. For the asynchronous conditions, the translation or expansion was in anti-phase for the two eyes. In the perturbed speeds condition, the speed at which the cylinder rotated oscillated sinusoidally between 15º/s and 25º/s. This oscillation was shifted by a quarter phase between the two eyes, so that the velocities did not match (but the paths were not perturbed). The new conditions are illustrated in Figure 5.5. The conditions were tested in separate blocks and compared to performance in Experiment 1.

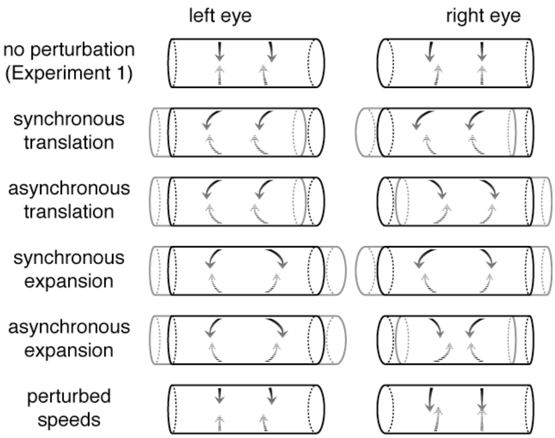

Figure 5.5: Schematic description of the perturbations in Experiment 2. Arrows show how dots at four positions on the cylinder move when the near side of the cylinder moves downwards. Two of the positions are aligned horizontally with the left eye and two with the right eye. Solid arrows show motion on the near side of the cylinder and dashed arrows show motion on the far side. The unperturbed condition of the first experiment is shown at the top for comparison. In all panels the (identical) initial condition is shown in black. If the cylinder's position or size changes, the position and size some time later is shown in grey. For clarity, the rotation of the cylinder is exaggerated, only the first quarter cycle of the perturbation is shown, and motion is shown for dots at the same positions in both eyes. Actual details of the perturbations are presented in the text.

Results

Monocular performance was generally slightly worse when the cylinder moved or expanded laterally, than it was without such perturbations (see Figure 5.6). The fact that even expansion did not reduce monocular performance to chance indicates that subjects must be able to pick up quite subtle complex patterns of relative motion because expansion changes the motion signals differently for different dots.

A more important finding is that for translation and varying speeds, performance with two eyes was better than with one. This was even so, though perhaps less so, when the translation was in anti-phase for the two eyes. Only adding expansion reduced binocular performance to the same level as monocular performance.

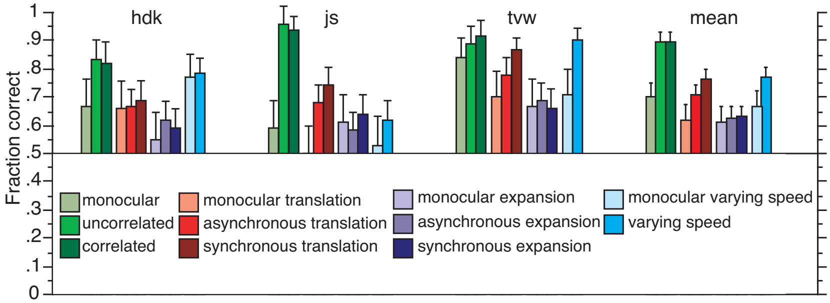

Figure 5.6: Individual and average performance of the three stereoblind subjects in Experiment 2. The three leftmost (green) bars of each set show performance in Experiment 1 for comparison.

Discussion

The stereoblind subjects' use of binocular information relies on something that is disrupted by changing the dots' relative paths (i.e. affecting individual dots' paths differently) in each eye, but not (or less so) by changing the dots' overall paths or their velocities along those paths, even if such changes differed between the eyes.

The reasonably good performance of the stereoblind subjects in our asynchronous translation condition (about 70% correct) is inconsistent with the idea of interocular velocity differences as described in Brooks & Stone (2006), Rokers et al. (2008) and Fernandez & Farell (2006). In this condition, points at corresponding positions in the two eyes always moved in opposite directions. According to the mechanisms proposed by Brooks & Stone (2006), Rokers et al. (2008) and Fernandez & Farell (2006), subjects should have perceived all dots to either be moving towards them or away from them, and thus not have been able to determine in which direction the cylinder is rotating. Given the subjects' reasonably good performance in this condition, this is clearly not the case.

Adding horizontal expansion to the rotating cylinder changes the relative direction of motion of each dot differently. This perturbation resulted in a drop in performance to the level achieved with one eye. This suggests that the stereoblind subjects use interocular differences in relative direction of motion to judge motion in depth, in a similar way that subjects with normal binocular vision use changes in relative disparity.

General Discussion

We found that all three stereoblind subjects due to eye patch treatment for amblyopia can use binocular information to determine the direction in which a transparent virtual cylinder is rotating. We suggest that they use differences between the eyes in the relative direction of motion of corresponding parts of the stimulus to do so. Having established that subjects do not simply rely on the dots' paths (because they can cope with translation), or velocity profiles (because they can cope with varying speeds), and since the local increase in distance between points that are approaching (expansion) is a monocular cue, we conclude that the binocular information that is used in our stimulus is the difference between the relative direction of retinal motion of points on the front and back of the cylinder in the two eyes. In our cylinder, these surfaces are part of the same object, but normally stereoblind subjects presumably use other (static) objects as references, comparing the direction of motion relative to such references in the two eyes. When a surface's slant changes (as in the study mentioned in the Introduction), however, the comparison may also be between parts of a single object.

That is, stereoblind subjects do not simply use interocular velocity differences, but rather interocular differences in relative direction of motion to judge the direction of motion in depth.

Chapter 6: Summary and Discussion

Chapter 2 and 3

Do timing differences between cues affect their combination in one estimate?

Different cues for the same attribute have been found to be processed independently from one another in different parts of the brain (Schmolesky et al., 1998). This independence of processing might result in timing differences between the cues' estimates if one cue takes longer to process than another. How does the brain handle such timing differences when integrating different cues? I examined this in Chapter 2.

In a virtual environment, subjects saw a virtual ring whose slant was indicated by either monocular cues, binocular disparity, or both. The ring jittered in one plane and the slant of this plane could change. Subjects were instructed to respond whenever they saw such a slant change (experiment 1), or, (if perceiving a slant change,) to indicate in what direction the ring changed slant (experiment 2). I introduced artificial timing differences between the changes in the two cues of up to 400 ms and studied how this affected subjects' performance. If the temporal asynchrony between the cues was smaller than about 100 ms, then subjects performed better with both cues available than expected by probability summation of their performances with only the individual cues. Larger asynchronies resulted in a decrease in performance to the level of probability summation. The findings suggest that the temporal resolution of neural processing will be so poor that any possible latency differences between the cues, which are expected to be in the order of tens of milliseconds, are likely to not be picked up and will thus probably be ignored when the cues are integrated into a single estimate of slant.

For the online control of movement it is important to respond quickly. If timing differences between cues are ignored, the extent to which cues will be effective in guiding our actions will depend on how fast they provide new information. If cues are always integrated, one may hypothesize that fast responses are usually initiated based on the information that is provided by the fastest cue. I tested this in Chapter 3. Whilst sitting in a virtual environment, subjects were instructed to put a cylinder on surface. The slant of this surface was specified by monocular cues, such as shape and texture of the surface, and its binocular disparity. A computer-controlled, real table was rotated to agree with the virtual surface. Immediately after movement onset, the virtual surface could change slant and subjects had to correct their movement so that the cylinder was in the proper angle before making contact with the table. Monocular cues and binocular disparity indicated the change at the same time or with a 150 ms timing difference. The slant change itself could be masked by a blank image during the time in which the virtual surface moved from one slant to another. As expected, subjects adjusted their movement to accommodate the changed surface slant. The onset of their correction depended on whether monocular or binocular cues changed first. Responses were fastest when the monocular cues changed first and the change was not masked. Masking the slant change delayed the subjects' responses for changes in which the monocular cues changed first, but not for changes in which binocular disparity changed first. This suggests that subjects could use three distinct types of information to base their response upon: the new slant as indicated by the static monocular images, the new slant as indicated by the disparity between the two eyes' two images and the slant change (the transient) in the monocular images. Responses to the change in the monocular images were approximately 40 ms faster than responses to a new binocular slant, and about 90 ms faster than responses to a new monocular slant. So the planning and control of fast responses are indeed based on the information provided by the fastest cue. In view of the delays that we measured, the adjustments of ongoing movement to changes in orientation will usually be initiated by the change in the monocular images of the stimulus. As such a change might not be informative about the surface slant itself (making it difficult to calculate in which position the movement needs to end) and is only available for a brief period of time, it is likely that later parts of the response might also be based on the new binocular and monocular slant estimates that become available during the movement.

Chapter 3 shows that latency differences between cues are visible in subjects' online control of movement. This strongly suggests that any timing differences between cues are indeed ignored. It thus confirms the interpretation of the data in Chapter 2. Additionally, Chapter 3 also gives an explanation for the asymmetry in performance seen over the different cue asynchronies in Chapter 2. The jitter of the ring in Chapter 2 would have masked slant information from the change in the monocular images. As information about a new binocular slant seems to be processed 50 ms faster than information from a new monocular slant, the actual difference in neural latency between the cues will be smaller for negative than for the same positive asynchronies (remember that asynchrony was expressed as the monocular cues changing *x* s after binocular disparity), explaining why subjects benefit longer from cue combination for negative than for positive asynchronies.

Chapter 4:

Is the precision of a cue the sole determinant of its weight?

As discussed in the *Introduction* of this thesis, an integration of cues can be considered optimal if it maximizes the precision of the combined estimate. This can be achieved if the estimates of the individual cues are summed linearly with weights that reflect their precision; the less variable the estimate, the higher its weight in the combined estimate. Previous studies found that subjects increase the weight of a cue if auditory or haptic feedback indicates that its estimate is correct, and decrease the weight if it is indicated to be incorrect (Atkins et al., 2001; Ernst et al., 2000; Jacobs & Fine, 1999). In these studies, however, the physical precision of the different cues was not manipulated. This suggests that the change in weights was brought about by the change in the perceived accuracy ("correctness") of the cues, which would be a violation of optimal cue combination theory.

Interestingly, Knill (2007) found that subjects change the weights that they give to different cues if they have evidence that one of the priors that they normally use to interpret (one of) the cues is incorrect. He found that when subjects became aware of a heightened occurrence of ellipses versus circles in a context, as evident from large conflicts between stereoscopic cues and foreshortening, subjects decreased the weight that they gave to foreshortening and increased the weight that they gave to disparity. He argued that the conflict between the cues caused subjects to interpret the shape of the stimulus as being more ambiguous, and this broadening of shape expectations decreased the precision of foreshortening. This decreased precision of foreshortening triggered the change in weights.

Could the change in weights that was found in the feedback studies (Atkins et al., 2001; Ernst et al., 2000; Jacobs & Fine, 1999) be related to a change in the judged precision of the cues? All these studies used conflicts between the cues to study the weights that subjects gave to each of the cues. This might have made subjects' interpretations about shape more ambiguous. If so, the change in weights should be slow (subjects first have to accumulate evidence that the shape is ambiguous) and independent from the feedback (thus always in favor of cues that do not depend on shape). In Ernst et al. (2000), the weight given to binocular disparity decreased when the feedback indicated that its estimate was wrong. If the change in weights was solely due to a change in shape expectations, disparity should always have been given more weight, regardless of the feedback. So it is likely that a change in people's shape expectations cannot fully explain the change in weights seen in these studies.

In Chapter 4, I tested whether the change in weights seen in feedback experiments is related to a change in the perceived precision or in the perceived correctness of the cues. Subjects were asked to put a cylinder on a slanted virtual surface. Surface slant was again defined by monocular cues and binocular disparity. On most trials, the two cues indicated conflicting slants. A real table (the one that was also used in Chapter 3) was rotated to match the slant of one of the cues. So at the end of their movement subjects received feedback about the correctness of each cue. On trials in which the two cues were in conflict, the shape of the virtual surface was a deformed trapezoid. On trials in which the two cues indicated the same slant, the surface's shape was square. If subjects noticed this difference and changed their shape expectations accordingly, the weight given to the monocular cues should decrease even when the feedback indicated that this cue was correct. The change in weights that was seen in response to a change in shape prior in Knill's study was slow. To test whether the change in weights in response to feedback studies might be caused by a change in the perceived precision of the shape cues (as subjects might start to notice the different shape of the surface on the conflict trials), I measured the time course of the change in weights and its dependence on the shape of the surface on former trials. To avoid interference of variables other than the feedback of the table (such as random variations in weights between blocks on different days), I introduced the switch in the feedback within the same session. Subjects first performed a small set of trials, in which the table indicated the same slant as one of the cues, and then a larger set of trials in which the table indicated the same slant as the other cue.

Subjects quickly gave more weight to the cue that was consistent with the feedback of the table. This change in weights occurred within just 12 trials and was clearly different for the two types of feedback. Its fast rate and its dependence on feedback suggests that it is unlikely that the change in weights resulted from a change in shape expectations. However, the difference in the shape of the stimulus between conflict and non-conflict trials might have been more 'obvious' in my experiment than in Knill's (2007), which might have increased the rate of the change in weights. To investigate this possibility, I tested whether the mean weight given to the monocular cues on the last 72 trials of the long block was different for trials that were preceded by conflict trials than for trials that were preceded by non-conflict trials. There was no difference, suggesting that the change in weights is not due to a change in subjects' shape expectations. Might subject have recalibrated the cues in response to the feedback? That is, did the feedback stimulate them to reinterpret the values of the cues. This would also manifest itself as a change in the weights in our analysis. However, recalibration should have affected responses on the non-conflict trials as well. The responses on the non-conflict trials were constant throughout the experiment, making it unlikely that recalibration caused change in weights.

Might the change in weights be related to a change in response strategy? Subjects might take the correctness of their percepts as felt on previous trials into account when perceiving the slant in new trials. For example, at the start of the experiment their percept of slant might have felt right, but after the change in feedback, they might have deduced that something had changed because their percepts were now clearly wrong. One obvious solution would be to simply go with the slant opposite to the slant that they are currently seeing (this is possible in our experiment because we have only two slants). If subjects used such a strategy, their responses over the entire experiment should be bimodally distributed. However, subjects' responses on the conflict trials were clearly unimodal in distribution. So the change in weight cannot be explained by a change in response strategy.

The fact that subject quickly change the weights that they give to binocular and monocular cues in response to changes in their perceived bias, without a change in the percepts on non-conflict trials, suggests that optimal cue combination theory needs to be extended to include subjects' judgments of a cue's correctness as a determinant of its weight. Thus Equation 2 of the *Introduction* needs to be altered to include a term in which the (past) deviation of a cue's value from that of other cues is captured. Additionally, the fast time constant of the change in weights suggests that subjects might have access to the estimates of the individual cues. When they notice from the felt slant at the end of their movement that their percept was wrong, a fast recovery as seen in our study can be accomplished by simply comparing the values of the individual cues to discover which one deviates and is thus likely to be wrong.

People can learn to use prior information to enhance their performance based on the likelihood that a situation occurs. For example, Körding & Wolpert (2004) found that subjects could learn to consider a probability distribution of likely positions when pointing towards positions in space, and more importantly, that subjects relied more on this prior when their sensory input was uncertain. In my experiment, the likelihood of encountering the two slants was equal in either cue. So the probability of the surface being in a certain slant did not change after the change in feedback. As a consequence, any influence of a slant prior on the combined estimate should have remained constant throughout the experiment.

I did manipulate the likelihood that one of the cues was correct. Subjects might have translated this into a 'correctness' prior that influences the perceived precision of the cue that indicated the same slant as the feedback, in a similar way as the shape prior in Knill's study (2007) influenced the perceived precision of foreshortening. Note that for such a posthoc evaluation to work, subjects need to know what cue indicated what value after their combination into one estimate. The change in weights then results from a change in this ‘correctness’ prior after the change in feedback. I think it is unlikely that this is the case because development of such a prior should have affected the end orientations in the non-conflict trials as well, as subjects should not have been able to distinguish between non-conflict and conflict trials. Since the estimates of the individual cues are likely to be biased, a small difference in the end orientation should be seen for non-conflict trials in which the feedback indicated that binocular disparity was correct and non-conflict trials in which the feedback indicated that the monocular cues were correct. Such a difference is not apparent; the lines seem to lie on top of each other.

One worrying possibility is that subjects *can* distinguish between conflict and non-conflict trials, and developed different priors for them. However, this would not only weaken the findings of the current experiment, but also that of other studies that used the cue conflict paradigm to study optimal cue combination. This paradigm calculates the weights of biased cues in conflict situations using subjects’ responses in non-conflict situations. This calculation is based on the premise that a cue’s perceived precision is not affected by the conflict. It might be that this assumption is wrong, and subjects are able to distinguish between conflict and non-conflict trials. If so, they then can develop priors independently for these two sorts of trials. If the correctness prior then influences the perceived precision of the cue differently for conflict and non-conflict trials, any calculations that are based on measurements in non-conflict situations will clearly be wrong. Not only in our study, but also in other studies in which feedback or influence of some sort of prior on the combined estimate might be expected.

Chapter 5:

What information do stereoblind subjects use to determine direction of motion in depth?

Surprisingly, when a stereoblind subject JS performed the experiment described in Chapter 3, he could adequately adapt his ongoing movement to changes in binocularly defined slant. Performing the experiment monocularly led to a failure to react, indicating that his responses were really based on binocular information. Interestingly, JS could respond when the movement of the virtual surface (from one slant to the next) was visible but not when it was masked. This suggests that JS used some sort of difference between the motion of the surface in the two eyes to determine in what direction the surface was moving. What can this binocular information be? Previous studies have shown that besides changing disparity and monocular motion signals, normally stereo sighted subjects can also use interocular velocity differences (IOVDs) to judge motion in depth (Brooks & Stone, 2006; Rokers *et al.*, 2008). The difference in the direction and speed of motions of the same stimulus between the two eyes is then used to estimate direction and speed of the stimulus moving in depth. Strabismus patients suffering from low stereo-acuity have been reported to be able to use this cue, too (Kitaoji & Toyama, 1987; Meada., 1999), suggesting that point-to-point correspondence between the eyes is not necessary for use of this cue.

In Chapter 5 of my thesis, I studied what exact binocular information three stereoblind subjects can use to judge rotational motion in depth. I presented JS, 2 other stereoblind subjects (HdK and TvW) and seven control subjects with normal binocular vision with a transparent virtual cylinder that rotated around its horizontal axis. The cylinder was defined by random dots that moved coherently towards or from their body for periods of 250 ms. Subjects were asked to indicate the direction of its rotation. The images of the cylinder that were presented to the two eyes were either correlated (taking the subjects' interocular distance into account) or uncorrelated (dots at a random different position for one eye than for the other). Uncorrelating the two eyes' images is expected to degrade information from changing disparity, but not necessarily information from interocular velocity differences, since for this last cue point-to-point correspondence in the two eyes might not be required. I used a perspective rather than an orthographic projection of the cylinder, so that in addition to changing disparity and interocular velocity differences, subjects could also use monocular motion signals to determine the direction of the rotation. I measured both monocular and binocular performance in the correlated condition and used subjects' monocular performance as a baseline to compare all their other performances to. The stereoblind performed better binocularly than monocularly, but more interestingly, they performed with uncorrelated images as well as with correlated images. While the control subjects all scored almost 100% in the correlated condition, none of them was able to perform better than their monocular performance in the uncorrelated condition. These findings suggest that the stereoblind subjects indeed use some difference in motion between the eyes to determine the direction of motion in depth. By contrast, the healthy controls do not seem to be able to extract this information.

In a second experiment, I isolated what exact difference in motion between the two eyes the three stereoblind subjects used to determine the direction of the rotation. I perturbed the motion of the dots in several ways and studied how this affected subjects' uncorrelated performance. In one condition, I added horizontal translation to the rotating cylinder. In another condition, I added horizontal expansion, making the cylinder 10% wider or narrower over time. Adding translation perturbs the relative direction of each dot on the cylinder's surface in a coherent way, whereas adding expansion perturbs the motion of each dot differently. The added movement could be in the same direction for the two eyes (synchronous) or in different directions (asynchronous). Finally, I perturbed the speed with which the cylinder rotated in the two eyes, so that the cylinder moved with a different speed in one eye than in the other. Adding translation or perturbing the speed of the cylinder lowered the stereoblinds' performance, but binocular performance was still better then monocular performance and performance in the synchronous translation condition was better than in the asynchronous translation condition. Put differently, in the translation and perturbed speeds conditions, the binocular information that they were using was still present, although less reliable. Adding expansion to the rotating cylinder lowered performance. More importantly, binocular performance was not better than monocular performance and there was no difference between synchronous and asynchronous performance. Thus, adding expansion to the cylinder disrupted the binocular information that the three stereoblind subjects used to determine the direction of the rotation. This suggests that they used the difference in relative direction of motion between the two eyes of the corresponding sets of points on one side of the cylinder to judge the direction in which it was rotating.

Why could the stereoblind subjects use such interocular differences in the relative direction of motion to determine the direction in which the cylinder was rotating in the uncorrelated condition, but subjects with normal binocular vision could not? The controls might simply not use this cue because information from changing disparity and monocular motion signals will normally be sufficiently reliable. That is, information from IOVDs might receive relatively less weight because subjects might judge the estimates of the monocular motion signals and changing disparity to be far more reliable. Another possibility is that changing disparity and interocular differences in motion share processing in people with normal binocular vision. For example, changing disparity might be used to determine corresponding motion in the two eyes for calculation of IOVDs. As changing disparity gives unreliable depth estimates in the uncorrelated condition (the random different positioning of dots in the left and right eye is likely to result in a high number of mismatches), the calculation of interocular differences in motion will be based on a faulty comparison of velocities between the eyes and thus be unreliable. The stereoblind subjects do not have access to information from changing disparity and thus have to use another source of information in order to determine which velocities correspond in the two eyes. For our stimuli, they use the relative 2-D direction of the motion for this purpose. So although the stereoblind subjects are not able to generate a classical stereo depth map based on disparity, they can use disparity for calculation of motion in depth.

General Conclusion and Discussion

Chapter 2 showed that artificially induced timing differences between slant cues of smaller than 100 ms do not significantly affect the quality of the combined estimate. Chapter 3 showed that latency differences between slant cues are visible in the online control of movement. The findings strongly suggest that the small latency differences between slant cues are simply ignored when the brain integrates them in one estimate of slant. As a consequence, in a dynamic context the combined estimate will evolve over time, initially reflecting information from only the fastest processed cue, but later changing as cues with a longer latency gradually gain influence.

Chapter 4 showed that subjects can quickly change the weight that they give to a cue in response to changed feedback about its correctness, without there being a change in its perceived precision. The change in weight could not be explained by a recalibration of the cues in response to the change in feedback or by a change in response strategy. This suggests that not only the variability of a cue, but also its systematic bias, can influence the weight that it receives in the combinational rule. The fast time constant of the change in weights suggests that next to the precision of each cue, subject might also have access to its value, so that they can quickly compare the estimates of the individual cues to discover which estimate is biased and change its cue's weight accordingly.

Chapter 5 showed that stereoblind subjects can use interocular velocity differences (IOVDs) to determine the direction of motion in depth. Uncorrelating the two eyes' images of a transparent rotating cylinder decreased performance to chance level or to monocular performance for subjects with normal binocular vision, but did not decrease performance of stereoblind subjects. These findings suggest that whereas subjects with normal stereovision cannot ignore erroneous estimates from changing disparity (indicating that their use of IOVDs is based on a point-to-point correspondence between the eyes, or they simply give very low weight to this IOVDs in the presence of other motion-in-depth cues), stereoblind subjects are not hindered by changing disparity and can reliably use the differences in motion between the eyes to deduce the direction of the rotation. Interestingly, the stereoblind subjects' performance fell back to chance level or to monocular performance when expansion was added to the rotational movement. So for calculation of motion in depth, they can correspond velocities between the eyes based on similar relative direction of movement.

How do the findings of the different chapters relate to each other? The fact that we find in Chapter 5 that stereoblind subjects can use disparity to calculate motion in depth but not a static stereo depth map (displaying 3D position), and in Chapter 3 that subjects with normal binocular vision can use both the motion and the new static position of the virtual surface to get an indication of its new slant, suggests that motion and position are processed separately by the brain when calculating depth.

In Chapter 3, subjects' responses to a change in binocularly defined slant were unaffected when the motion of the surface was masked. I concluded that this was because subjects only responded to binocular slant, i.e. the new slant that was indicated by the relative disparities after the surface had stopped moving, and not to the changing disparity of the surface during the period in which it moved from one slant to the next. In Chapter 5, however, some of these subjects could readily discriminated between directions of a continuously rotating virtual cylinder. Why did the subjects in Chapter 3 not use the changing disparity to alter their movements? An explanation could be that changing disparity has a longer latency than static disparity, maybe because changing disparity is simply the change in static disparities over time. But then, why is the change in the monocular images faster than monocularly defined slant? For the stimulus used in Chapter 3, subjects do not only have access to the change in monocularly defined slant over time, but also to estimates about the direction in which the surface is changing from the expansion and contraction of texture elements on the top and bottom of the surface. The change in the monocular image that subjects were responding to could simply have been this expansion and contraction on the top and bottom of the surface, which might be processed faster than changes in monocular defined slant.

I would like to conclude with the notion that the most important thing that I learned from the experiments that I discussed in this thesis is that we need to consider the whole dynamics of a context in detail, as well as the dynamics of the mechanisms in the brain that reconstruct this context for subsequent action, before we can evaluate whether people make statistically optimal use of all available information.

References

Adams, W. J., & Mamassian, P. (2004). Bayesian combination of ambiguous shape cues. *Journal of Vision, 4*(10), 921-929.

Arnold, D. H., & Clifford, C. W. (2002). Determinants of asynchronous processing in vision. *Proceedings of the Royal Society London B, 269*(1491), 579-583.

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research, 41*(4), 449-461.

Aymoz, C., & Viviani, P. (2004). Perceptual asynchronies for biological and non-biological visual events. *Vision Research, 44*(13), 1547-1563.

Bartels, A., & Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Research,* 46, 2280 - 2286.

Brenner, E., & Smeets, J. B. J. (2000). Comparing extra-retinal information about distance and direction. *Vision Research, 40*(13), 1649-1651.

Brenner, E., & Smeets, J. B. J. (2006). Two eyes in action. *Experimental Brain Research, 170*(3), 302-311.

Brenner, E., van den Berg, A. V., & van Damme, W. J. (1996). Perceived motion in depth. *Vision Research,* 36, 699-706.

Bresciani, J. P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: Auditory signals can modulate tactile tap perception. *Experimental Brain Research, 162*(2), 172-180.

Bridge, H., & Parker, A. J. (2007). Topographical representation of binocular depth in the human visual cortex using fmri. *Journal of Vision, 7*(14), 15 11-14.

Brooks, K. R., & Stone, L. S. (2006). Spatial scale of stereomotion speed processing. *Journal of Vision, 6*(11), 1257-1266.

Chandrasekaran, C., Canon, V., J.C., D., Kourtzi, Z., & Welchman, A. E. (2007). Neural correlates of disparity-defined shape discrimination in the human brain. *Journal of Neurophysiology, 97*(2), 1553-1565.

Cumming, B. G., & Parker, A. J. (1994). Binocular mechanisms for detecting motion-in-depth. *Vision Research, 34*(4), 483-495.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

Ernst, M. O., Banks, M. S., & Bulthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience, 3*(1), 69-73.

Fernandez, J. M., & Farell, B. (2006). Motion in depth from interocular velocity differences revealed by differential motion aftereffect. *Vision Research, 46*(8-9), 1307-1317.

Greenwald, H. S., Knill, D. C., & Saunders, J. A. (2005). Integrating visual cues for motor control: A matter of time. *Vision Research, 45*(15), 1975-1989.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science, 298*(5598), 1627-1630.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12), 967-992.

Hogervorst, M. A., & Brenner, E. (2004). Combining cues while avoiding perceptual conflicts. *Perception, 33*(10), 1155-1172.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research, 39*(21), 3621-3629.

Jacobs, R. A., & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research, 39*(24), 4062-4075.

Kitaoji, H., & Toyama, K. (1987). Preservation of position and motion stereopsis in strabismic subjects. *Investigative Ophthalmology Vision Science, 28*(8), 1260-1267.

Knill, D. C. (1998). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research, 38*(11), 1683-1711.

Knill, D. C. (2007). Learning bayesian priors for depth perception. *Journal of Vision, 7(8)*(13), 1-20.

Knill, D. C. & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43*(24), 2539-2558.

Körding, K.P. & Wolpert, D.M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427(1501),* 244-247

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience, 20*(9), 3310-3318.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35*(3), 389-412.

Louw, S., Smeets, J. B. J., & Brenner, E. (2007). Judging surface slant for placing objects: A role for motion parallax. *Experimental Brain Research*, *183*(2), 149-158.

Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*(11), 1432-1438

Maeda, M., Sato, M., Ohmura, T., Miyazaki, Y., Wang, A. H., & Awaya, S. (1999). Binocular depth-from-motion in infantile and late-onset esotropia patients with poor stereopsis. *Investigative Ophthalmology Vision Science, 40*(12), 3031-3036.

McKee, S. P., Levi, D. M., & Movshon, J. A. (2003). The pattern of visual deficits in amblyopia. *Journal of Vision, 3*(5), 380-405.

Moutoussis, K., & Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society London B Biol Sci, 264*(1380), 393-399.

Moutoussis, K., & Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society London B Biol Sci, 264*(1387), 1407-1414.

Muller, C. M. P., Brenner, E., & Smeets, J. B. J. (2007). Living up to optimal expectations. *Journal of Vision, 7*(3), 2.

Muller, C. M. P., Brenner, E., & Smeets, J. B. J. (2008). Testing a counter-intuitive prediction of optimal cue combination. *Vision Research,* 49, 134-139.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the mcgurk effect. *Perception & Psychophysics, 58*(3), 351-362.

Nishida, S., & Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology, 12*(5), 359-368.

Oruc, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*(23), 2451-2468.

Poom, L. (2002). Visual binding of luminance, motion and disparity edges. *Vision Research, 42*(23), 2577-2591.

Rokers, B., Cormack, L. K., & Huk, A. C. (2008). Strong percepts of motion through depth without strong percepts of position in depth. *Journal of Vision, 8*(4), 6 1-10.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., et al. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology, 79*(6), 3272-3278.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*(1), 147-152.

Smeets, J. B. J., van den Dobbelsteen, J. J., de Grave, D. D., van Beers, R. J., & Brenner, E. (2006). Sensory integration does not lead to

sensory calibration. *Proceedings of the National Academy of Sciences of the United States of America, 103*(49), 18781-18786.
Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-d shape from binocular disparity and structure from motion. *Perception & Psychophysics, 54*(2), 157-169.
van Beers, R. J., Sittig, A. C., & Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology, 81*(3), 1355-1364.
van Ee, R., Adams, W. J., & Mamassian, P. (2003). Bayesian modeling of cue interaction: Bistability in stereoscopic slant perception. *Journal of the Optical Society of America. A, Optics, Image Science and Vision, 20*(7), 1398-1406.
van Mierlo, C. M., Brenner, E., & Smeets, J. B. J. (2007). Temporal aspects of cue combination. *Journal of Vision, 7(7)*(8), 1-11.
Van Mierlo, C. M., Brenner, E., & Smeets, J. B. J. (2009). Slant cues are processed with different latencies for the online control of movement. *Journal of Vision, 9(3)*(25), 1-8.
Viviani, P., & Aymoz, C. (2001). Colour, form, and movement are not perceived simultaneously. *Vision Research, 41*(22), 2909-2918.
Wu, D. A., Kanai, R., & Shimojo, S. (2004). Vision: Steady-state misbinding of colour and motion. *Nature, 429*(6989), 262.
Wuerger, S. M., Hofbauer, M., & Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics, 65*(8), 1188-1196.
Yuille, A., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. *Perception as Bayesian Inference,* 123-161.
Zeki, S., & Moutoussis, K. (1997). Temporal hierarchy of the visual perceptive systems in the mondrian world. *Proceedings of the Royal Society of London B Biol Sci, 264*(1387), 1415-1419.

Samenvatting en Conclusie

Cue combinatie voor het inschatten van helling in perceptie en actie

Introductie

Om een beker op kunnen pakken hebben we verschillende soorten informatie nodig. Om te zorgen dat onze arm en hand de juiste afstand overbruggen en in de juiste oriëntatie eindigen is het belangrijk dat we de 3D posities van deze lichaamsdelen en de beker weten. Deze posities kunnen we visueel inschatten aan de hand van verschillende monoculaire informatiebronnen die aanwezig zijn in het 2D retinale beeld, zoals de grootte, vorm en relatieve positie van de projectie van het object. De projectie van de beker op de retina zal kleiner zijn en hoger in het totale beeld staan als de beker verder weg staat. De vorm van de projectie hangt af van de oriëntatie van de beker. De grootte van de textuurelementen en de gradiënt is wederom afhankelijk van de exacte oriëntatie en afstand van het object. Naast dit soort monoculaire informatie kunnen we ook diepte informatie halen uit het verschil in beeld tussen de twee ogen. Omdat de twee ogen niet op dezelfde positie in het hoofd zitten zal de projectie van licht in het enige oog anders zijn dan in het andere oog. Dit verschil, de binoculaire dispariteit, is groter wanneer een object dichtbij ons lichaam staat dan wanneer het verder weg staat, en dispariteit geeft dus ook diepte informatie.

Verschillende onderzoeken hebben gevonden dat mensen eerst de monoculaire and binoculaire informatie afzonderlijk van elkaar verwerken in aparte gedeelten van de hersenen, en daarna hun inschattingen van diepte optimaal integreren. De combinatie wordt optimaal genoemd omdat het brein rekening houdt met de hoeveelheid ruis of variabiliteit in ieder soort informatie. Informatie met veel ruis krijgt weinig gewicht, informatie met weinig ruis krijgt veel gewicht in de gecombineerde inschatting van diepte.

Timing

Zoals eerder gezegd, wordt de monoculaire en binoculaire informatie eerst apart verwerkt. Doordat de ene informatie misschien iets meer tijd nodig heeft om verwerkt te worden dan de andere, kunnen er tijdsverschillen tussen eenzelfde gebeurtenis in de twee soorten informatie ontstaan. In Hoofdstuk 2 en 3 heb ik onderzocht hoe tijdsverschillen tussen gebeurtenissen in monoculaire and binoculaire informatie van invloed zijn op hun combinatie in één inschatting van diepte. In Hoofdstuk 2, presenteerde ik proefpersonen met een in zijn eigen vlak rond bewegende ring, die zo nu en dan van hoek veranderde. De helling van de ring kon ingeschat worden aan de hand van de vorm van de ring en de textuur gradiënt van puntjes in de ring, en aan de hand van het verschil in beeld tussen de twee ogen. Proefpersonen moesten reageren zo gauw ze de ring van hoek zagen veranderen (Experiment 1) of aangeven in welke richting de hoek was veranderd (Experiment 2). De hoekverandering kon eerst in de binoculaire informatie gebeuren of eerst in de monoculaire informatie of in beide soorten informatie tegelijkertijd. In de asynchrone condities varieerde de tijd tussen de verandering in de monoculaire informatie en de verandering in de binoculaire informatie tussen 5 tot 400 milliseconden en of de monoculaire informatie kon eerst veranderen of de binoculaire informatie. Proefpersonen konden het best de hoekveranderingen detecteren, of tussen hoekveranderingen discrimineren, wanneer de binoculaire en monoculaire informatie tegelijkertijd veranderden. Voor de asynchrone veranderingen bleef de presentatie van de proefpersonen maximaal tot het timingsverschil op liep tot ongeveer 100 ms, waarna prestatie zakte naar de prestatie die de proefpersonen hadden met een verandering in mar één van de 2 soorten informatie (en de ander niet veranderde). De resultaten van dit experiment suggereren dat tijdsverschillen tussen binoculaire en monoculaire informatie van minder dan 100 ms geen invloed zullen hebben op de kwaliteit van de gecombineerde inschatting.

Als ze geen invloed hebben op de gecombineerde inschatting, wat gebeurt er dan met de kleine kleine tijdsverschillen? Worden ze genegeerd tijdens de integratie, of worden ze op een of andere manier gecompenseerd? Ik heb dit onderzocht in Hoofdstuk 3. Proefpersonen zaten in een virtuele omgeving waar ze een object op een oppervlak moesten zetten. Dit oppervlak kon van hoek veranderen tijdens hun beweging, en ze moesten daarop reageren door de eind oriëntatie van hun hand aan te passen. De verandering in hoek kon eerst in de binoculaire informatie, eerst in de monoculaire informatie, of in beide soorten informatie tegelijkertijd gebeuren. In reactie op de verandering in hoek, corrigeerden proefpersonen hun beweging en de start van deze correctie was afhankelijk van welke informatie er eerst veranderde. Correcties waren 90 ms sneller wanneer de monoculaire informatie eerst veranderde dan wanneer de binoculaire informatie eerst veranderde. Als we de verandering in monoculaire informatie maskeerden zodat de beweging van het oppervlak niet te zien was, reageerden mensen 50 ms later. Toen we de verandering in de binoculaire informatie maskeerden, reageerde mensen even snel als zonder deze maskering. Wanneer beide cues tegelijkertijd veranderden was de correctie even snel als die met de snelste cue. Het feit dat we verschillen in de start van correcties kunnen meten als we de timing van de verandering in de cues manipuleren suggereert dat de verschillen in verwerkingstijd niet gecompenseerd worden voor de integratie. Bovendien impliceert de data dat mensen in de monoculaire informatie zowel de verandering in oriëntatie (de transiënt) als de veranderde oriëntatie kunnen gebruiken om de helling in te schatten, maar alleen de veranderde oriëntatie in binoculaire informatie.

Precisie versus Bias

Als het brein een precies maar incorrect signaal combineert met een ruizig maar correct signaal, dan zal volgens de huidige theorie van optimale cue combinatie de precieze maar incorrecte informatie meer gewicht krijgen in de gecombineerde inschatting dan de ruizige correcte informatie. Dit zou zeer nadelig zijn voor interactie met het object.

In Hoofdstuk 4 heb ik onderzocht of het brein inderdaad werkelijk op deze wijze omgaat met de ingeschatte correctheid van informatie wanneer het twee signalen integreert in één inschatting van oriëntatie. Proefpersonen zaten wederom in een virtuele omgeving waarin ze object op een oppervlak moesten zetten. De helling van dit oppervlak kon worden afgeleid uit binoculaire and monoculair informatie, welke onafhankelijk van elkaar konden worden gemanipuleerd. De monoculaire and binoculaire informatie gaven allebei dezelfde oriëntatie aan, of conflicterende oriëntaties. Het virtuele oppervlak stond op dezelfde afstand als een echte tafel. Deze tafel kon roteren om zijn horizontal as, en werd in dezelfde oriëntatie gedraaid als aangegeven werd door of de monoculaire informatie of de binoculaire informatie, zodat proefpersonen aan het eind van hun beweging voelden of hun inschatting goed was of niet. De tafel gaf de eerste 20 trials de oriëntatie van de ene soort informatie en dan voor 100 trials de oriëntatie van de andere soort informatie aan. Dus de aangegeven accuraatheid van de monoculaire en binoculaire informatie veranderde na een korte tijd

Hoe reageren de proefpersonen hier op? Als alleen de precisie van belang is voor het gewicht van de informatie in de gecombineerde inschatting, dan zou de verandering in feedback geen invloed moeten hebben op de inschatting (aangezien we alleen de ingeschatte correctheid van de informatie manipuleren, niet de ruis). Als de correctheid van de informatie wel van invloed is op het gewicht van de informatie in de gecombineerde inschatting, dan zou hun inschatting van de hoek moeten veranderen na de verandering in feedback. De proefpersonen schatte de hoek van het oppervlak na de feedback verandering anders in. Dit was zichtbaar als een verandering van de eind oriëntatie van hun hand na de verandering feedback. Dit gebeurde alleen op trials waar de monoculaire en binoculaire informatie in conflict waren, niet als ze niet in conflict waren.

Dit betekent dat de verandering in de proefpersoon’s inschatting te wijten was aan een verandering in het gewicht, niet aan een herinterpretatie van de informatie (want dan zou de inschatting op non-conflict trials ook veranderd moeten zijn). Er was geen verschil in de eind oriëntatie van de hand tussen trials die waren vooraf gegaan door conflict trials en trials die vooraf waren gegaan door non-conflict trials (de vorm van de virtuele tafel was iets anders in de conflict trials, wat voor verwarring had kunnen zorgen). Dus de ervaren precisie (ruis) van de monoculaire en binoculaire informatie was constant. Tenslotte, waren de eind oriëntaties unimodaal verdeeld voor alle conflicten. Als de verandering in feedback een verandering in response bias had veroorzaakt dan hadden we bimodale distributies moeten zien. De resultaten en controles van dit experiment suggereren dat het brein wel degelijk rekening houdt met de correctheid van de verschillende soorten informatie wanneer het ze integreert in één inschatting. Het zal incorrecte informatie minder gewicht geven dan correcte informatie, ook al is er geen verschil in precisie.

Beweging in diepte

Niet iedereen kan dispariteit gebruiken om diepte in te schatten. Er zijn mensen die nooit geleerd hebben om te twee ogen te corresponderen omdat de retinale beelden te veel verschilden (bijvoorbeeld om dat er ooit wat mis met een van de twee ogen). Jeroen Smeets, mijn promotor, is door (behandeling van) een lui oog stereoblind, dat wil zeggen, hij kan niet identificeren welke positie in het linkeroog bij welke positie in het rechteroog hoort (gegeven dat ze hetzelfde object voorstellen). Hierdoor kan hij niet dispariteit gebruiken om diepte in te schatten.

Jeroen heeft in het verleden vaak meegedaan aan mijn proeven om te controleren of dat er tijdens mijn manipulaties van binoculaire dispariteit geen monoculaire artefacten aanwezig zijn die proefpersonen kunnen gebruiken in plaats van de door mij gemanipuleerde binoculaire informatie. Zeer verontrustend in eerste instantie, vond ik dat Jeroen adequaat kon reageren op de manipulatie in binoculaire dispariteit in Hoofdstuk 3. Echter, wanneer hij het experiment met één oog dicht uitvoerde, of wanneer we de beweging van de tafel (van de ene naar de andere helling) maskeerden, verdween zijn reactie. Dit suggereert dat hij niet een monoculair artefact gebruikte om te reageren, maar daadwerkelijk de dispariteiten die aanwezig zijn tijdens de beweging van het virtueel oppervlak van de ene naar de andere hoek. In Hoofdstuk 5 heb ik onderzocht wat deze binoculaire informatie is, of andere stereoblinde mensen deze informatie ook kunnen gebruiken en op welke manier JS en eventueel andere stereoblinde proefpersonen achterhalen welke objecten in het linkeroog ze moeten vergelijken met welke objecten in het rechteroog.

Voorgaand onderzoek heeft laten zien dat bij beweging in diepte mensen niet alleen verandering in dispariteit en verandering in monoculaire cues (zoals in grootte en vorm) kunnen gebruiken, maar ook het verschil in de bewegingen tussen de projecties van de objecten in de twee ogen. Bovendien zijn er aanwijzingen dat niet alleen mensen met normaal binoculair zicht dit kunnen, maar ook mensen met stereo anomalie of stereoblindheid. In Hoofdstuk 5, heb ik uitgezocht of stereoblinde proefpersonen inderdaad het verschil in beweging tussen de ogen kunnen gebruiken om veranderingen van oriëntatie te beoordelen, en zo ja, op basis van welke informatie zij dan de snelheden in de twee ogen corresponderen.

Drie stereoblinde proefpersonen en 7 controle proefpersonen met normaal binoculair zicht zagen een horizontaal transparante cilinder die in 2 mogelijke richtingen om zijn horizontale as kon roteren. De proefpersonen moesten aangeven of de cilinder naar hun toe draaide of van hun af. De transparante cilinder was gedefinieerd door kleine rode puntjes die voor korte tijd (250 ms) over het lichaam van de cilinder bewogen en dan ververst werden op een willekeurig andere locatie op de cilinder. Om gebruik van interoculaire verschillen in beweging onafhankelijk van binoculaire dispariteit te kunnen bestuderen, presenteerde ik een cilinder die door willekeurig andere puntjes werd gedefinieerd aan het enige oog dan aan het andere oog (dus de beelden van de cilinder waren ongecorreleerd). Dit interfereert met correspondentie voor binoculaire dispariteit maar niet perse met correspondentie voor interoculaire verschillen in beweging. Daarnaast meette ik ook prestatie met gecorreleerde beelden van de cilinder (het beeld voor het rechteroog was x mm verschoven ten opzichte van het beeld voor het linker oog, x afhankelijk van de afstand tussen de proefpersoon zijn ogen) en prestatie met één oog dicht (monoculair).

Een aantal van de stereoblinde en controle proefpersonen konden de richting van de rotatie bepalen aan de hand van kleine monoculaire bewegingssignalen in de stimulus. Ze rapporteerden dat ze de beweging van de puntjes volgden en dan aan de hand van de kromming van hun paden achterhaalden in welke richting de puntjes op de voorkant bewogen. Met gecorreleerde beelden presteerden de controle proefpersonen bijna 100% goed, maar met ongecorreleerde beelden presteerden ze op kansniveau of op hetzelfde niveau als hun monoculaire prestatie. De stereoblinde proefpersonen presteerden net zo goed met gecorreleerde als ongecorreleerde beelden: allebei rond 70% goed. Belangrijker, hun binoculaire prestatie was beter dan hun monoculaire prestatie. Deze resultaten suggereren dat de stereoblinde proefpersonen inderdaad bewegingsverschillen tussen de ogen kunnen gebruiken om de richting van beweging in diepte in te schatten.

In een tweede experiment onderzocht ik op basis van welke informatie de stereoblinde proefpersonen bewegingen in de twee ogen corresponderen. De drie stereoblinden zagen weer ongecorreleerde beelden van de cilinder. In één conditie roteerde de cilinder niet alleen, maar maakte hij ook een translatie naar links of een translatie naar rechts. De translatie was in dezelfde richting in beide ogen of in tegenovergestelde richting. In een andere conditie bewoog de cilinder in het ene oog met een andere snelheid dan in het andere oog. In een laatste conditie roteerde de cilinder niet alleen, maar expandeerde of contracteerde hij tegelijkertijd (wederom in dezelfde richting of in tegenovergestelde richting in de twee ogen). Toevoeging van een translatie of het manipuleren van de snelheden in de twee ogen verlaagde de ongecorreleerde prestatie van de stereoblinde proefpersonen, maar de prestatie was nog steeds hoger dan hun monoculaire prestatie of kansniveau. Dit betekent dat de binoculaire bewegingsinformatie nog steeds aanwezig was, alhoewel ik het de stereoblinde proefpersonen wel moeilijker had gemaakt om deze informatie te gebruiken. Toevoeging van expansie daarentegen verlaagde hun prestatie naar kansniveau of naar hun monoculaire prestatie, zowel voor expansie in dezelfde richting in de twee ogen als voor expansie in tegenovergestelde richting. Dit suggereert dat als we de verandering in richting voor ieder puntje in de twee ogen op andere wijze manipuleren, de stereoblinde proefpersonen de snelheden in de twee ogen niet meer kunnen corresponderen. Dus de stereoblinde proefpersonen gebruiken het interoculaire verschil in de relatieve richting van de beweging om de beweging in diepte in te schatten. Ze omzeilen zo hun ‘niet optimale’ correspondentie op basis van positie die normaliter wordt gebruikt voor het bepalen van binoculaire dispariteit.

Algemene conclusie

Ik heb laten zien dat verschillen in verwerkingstijd tussen monoculaire en binoculaire informatie niet van invloed zijn op hun combinatie in één inschatting en gewoon genegeerd worden tijdens hun integratie. Dit betekent dat in een dynamische omgeving, diepte inschattingen aanvankelijk gedomineerd zullen worden door monoculaire informatie en de invloed van binoculaire dispariteit pas later toeneemt. Daarnaast heb ik laten zien dat niet alleen de precisie van monoculaire en binoculaire informatie van invloed is op hun gewicht in de inschatting, maar ook hun correctheid. Wanneer proefpersonen denken dat één van de twee soorten informatie niet klopt, verminderen ze het gewicht van deze informatie in de volgende combinaties. Dit gebeurt binnen enkele trials, wat impliceert dat de individuele informatiebronnen ook nog na hun integratie beschikbaar blijven voor verdere evaluatie van de kwaliteit van de gecombineerde inschatting. Tenslotte heb ik met behulp van stereoblinde proefpersonen laten zien dat het brein interoculaire verschillen in beweging afzonderlijk van verschillen in positie kan gebruiken om verplaatsing in diepte in te schatten. De correspondentie van de twee ogen is dan gebaseerd op een karakteristiek van de beweging zelf, de relatieve richting, in plaats van op positie.

Dankwoord

Als eerste wil ik mijn moeder bedanken. Mam, ik ben je heel erg dankbaar dat je mij zoveel helpt, en dat je zelfs de eerste maanden mee bent gegaan naar Birmingham zodat voor Findlay de overgang naar de nieuwe crèche makkelijker is verlopen en ik mij volledig kon concentreren op mijn nieuwe baan. Zonder jou was dit proefschrift niet tot stand gekomen.

Dan wil ik natuurlijk mijn man, Rory, en mijn zoontje Findlay bedanken. Rory, zonder jouw steun en interesse in mijn werk was ik niet gekomen waar ik nu ben. Findlay, je verlicht mijn leven, en blijft een inspiratiebron voor nieuwe onderzoeksvragen.

Ook mijn schoonouders Ton en Felicity wil ik bedanken. Jullie hebben vaak ingesprongen toen Findlay ziek was en niet naar de creche kon, of wanneer Rory op veldwerk was. Dank je wel.

Eli en Jeroen, ik waardeer het ontzettend dat jullie altijd aanwezig waren, nooit te weinig tijd hadden om iets uit te leggen, of om met mij ergens in discussie over te gaan. Ik heb in de 4 jaar bij jullie ongelofelijk veel nuttige vaardigheden geleerd voor het verrichten van onderzoek. De flexibiliteit die jullie getoond hebben tijdens mijn zwangerschap, de geboorte en de eerste maanden met Findlay heb ik als zeer prettig ervaren. Ook de opvoedkundige adviezen waren zeer welkom! Ik hoop dat we in de toekomst veel blijven samenwerken.

Mijn collega's uit Amsterdam en Rotterdam; Hanneke, Chris, Krista, Marianne, Denise, Rob, Stefan, Jeroen, Bernadette, Femke en Willemijn. Dank jullie wel voor de gezellige thee & lunchpauzes, conferenties, workshops en summerschools. Een van de belangrijkste dingen aan werk is toch dat er leuke collega's zijn, en jullie hebben mijn tijd in Rotterdam en Amsterdam dan ook zeer opgevrolijkt.

Mijn broertje Willem en zusje Cecilia; bedankt voor de kritische vragen over het nut van mijn onderzoek. ☺

Ten slotte wil ik mijn vrienden, Pauline, Sander, Anton, Nicolien, Jan-Willem, Annelies, Marc, Lenneke, Joost, Gerben, Iris, Pieter, Eefje, Thijs, Linda, Maurits, Menneke, Janneke, Arnold (en ieder ander die ik vergeten ben) bedanken voor de af en toe broodnodige afleiding van de wetenschap.

Curriculum Vitae

I was born on 31 July 1979 in Nijmegen, the Netherlands. Two weeks after that, I moved to Zoetermeer, where I went to primary (‘the Elzenhoek’) and secondary school (Alfrink College).

In 1997, I started studying Pharmacy at the University of Utrecht. After almost two years, I decided that this field was not for me. I wanted to more know about the brain workings of people, not more about drugs. In 1999, I switched over to Psychology, where I got interested in Neuropsychology and related fields. In 2004, I graduated on a study of cross-modal (auditory-tactile) attention in which I used ERPs to localize and compare auditory, tactile and cross-modal processing between blind and normally sighted subjects.

In April 2004, I started my PhD with Eli Brenner and Jeroen Smeets. The topic was ‘The binding problem in perception and action’ and they specifically intended me to study ‘the effects of asynchrony’. We stuck to that for 2 experiments and then went on to explore more interesting issues in cue combination for slant and beyond.

Since January 2009, I work as a postdoc in Andrew Welchman’s Binocular Vision lab at the School of Psychology of the University of Birmingham. I am currently investigating how the binocularly determined depth of specular highlights influence people’s percepts of the underlying object’s shape.

In my spare time, I play and cuddle with my 2-year old son Findlay, have meaningful conversations (among other things) with my husband Rory, and explore the vast space of the UK with friends or family.

Publications

Van Mierlo, C.M., Brenner, E.B., Smeets, J.B.J. (2007) Temporal aspects of cue combination. *Journal of Vision, 7(7)*(8), 1-11

Van Mierlo, C.M., Brenner, E.B., Smeets, J.B.J., Louw, S., Slant cues are processed with different latencies for the online control of movement, *Journal of Vision, 9(3)*(25), 1-8

Van Mierlo, C.M., Brenner, E.B., Smeets, J.B.J., Reliability is more than precision: Fast adaptation of slant cues' weights based on their 'correctness', *in preparation*

Van Mierlo, C.M., Brenner, E.B., Smeets, J.B.J., Motion in depth from interocular differences in the relative direction of movement, *under review (Vision Research)*

Van der Lubbe, R.H.J., Van Mierlo, C.M., Postma, A., Secondary modality-specific involvement of posterior brain areas in early blind, *in progress*

Verstijnen, I.M., van Mierlo, C.M., de Ruijter, P. (2008). Detrimental effects of earphone conversation on auditory environmental monitoring of visually impaired people, *Behaviour & Information Technology*, *27*(6), 507-516